Causal Analysis of Hydrological Systems

Case study of the Lhomme Karst System, Belgium.

Damien Delforge

Université catholique de Louvain
Earth and Life Institute – Environmental Sciences
Louvain-la-Neuve (Belgium)

November 2020

Causal Analysis of Hydrological Systems

Case study of the Lhomme Karst System, Belgium.

Damien Delforge[1,2,3]

[1]UCLouvain, Earth and Life Institute

[2]Royal Observatory of Belgium (ROB)

[3]FRIA grantee of the Fonds de la Recherche Scientifique-FNRS (Belgium).

Keywords: hydrology, empirical time-series analysis, causality, understanding, complexity, systems theory, nonlinearity, perceptions, sensitivity analysis, streamflow recession analysis, low flow forecasting, electrical resistivity tomography, clustering, hydrological connectivity, causal inference methods.

Thesis presented with a view to obtaining the title of PhD in Agronomic Sciences and Biological Engineering.

Thèse présentée en vue de l'obtention du titre de Docteur en Sciences Agronomiques et Ingénierie Biologique.

Supervisors

Marnik Vanclooster (UCLouvain)

Michel Van Camp (ROB)

Thesis Advisory Committee

Marnik Vanclooster (UCLouvain)

Michel Van Camp (ROB)

Olivier Kaufmann (UMons)

Vincent Hallet (UNamur)

Thesis Jury Members

Marnik Vanclooster (UCLouvain)

Michel Van Camp (ROB)

Olivier Kaufmann (UMons)

Olivier de Viron (LIENSs, la Rochelle Université)

Michel Crucifix (UCLouvain)

Presided by:

Emmanuel Hanert (UCLouvain)

Abstract

What does it take to have a causal understanding of complex hydrological systems from empirical time-series analysis? This is the starting point of the thesis focused on causality, which begins by introducing the facets of complexity. Complexity is related to the difficulty of understanding, which can be linked to the analysts, their ability to observe the system, elucidate it among all potential realities. Complexity could also be a trait of systems themselves presenting high dimensionality, nonlinear mechanisms, and a lack of self-organization. Understanding, nonlinearity, dimensionality, and organization, are the four angles investigated in this thesis and covered respectively in chapters 2, 4, 5, and 6. The last three are hydrogeological study cases applied to Lhomme Karst System, Rochefort, Belgium (LKS, Chapter 3), given that karst systems are particularly heterogeneous and complex hydrogeological environments.

In Chapter 2, the question of understanding the hydrological system is addressed from a philosophical essay on the concept of causality. The history of causality is traced from Ancient Greece to the present day, highlighting the major philosophical issues related to the concept while comparing them to epistemological issues in hydrology. The thesis provides a unifying and philosophical framework to address causality in science. Chapter 4 studies how the LKS affects the nonlinear behavior of the Lhomme River recession dynamics. A parsimonious method borrowed from the theory of nonlinear dynamical system is adapted for the case of recession analysis. Karst-induced nonlinearity is successfully detected in the LKS, while conventional methods of nonlinear recession analysis in hydrology do not reveal such an increase in complexity. Chapter 5 is focused on the dimension reduction of a time-lapse Electrical Resistivity Tomography (ERT) model of the subsurface above the Rochefort cave, using Time-Series Clustering (TSC). Using various time-series representations, resistivity series are clustered to extract spatiotemporal temporal patterns using three different algorithms. Similarly, four causal inference methods, which reports the organization of time-series variables as a causal graph are compared in Chapter 6. Three different study cases are considered, among which a virtual experiment, a case aiming at detecting preferential flow path from drip discharge data within the cave and ERT spatiotemporal patterns, a case framing the systemic behavior of the Rochefort cave system using relative gravimetry data that monitor changes in the mass

balance. Both Chapter 5 and 6 bring methodological awareness respectively on the TSC algorithms and causal inference methods and the variability and robustness of the reported results. Their conclusions stress the importance of conducting sensitivity analysis in both cases and that confidence in the results comes from redundant patterns while varying the approaches.

The conclusion makes consistent links between the philosophical framework and the potentialities and difficulties experienced in each chapter. It is concluded that what is needed to understand a complex system goes beyond a causal inference algorithm. It starts upstream with data, requires resources, questioning and philosophy, clear objectives, rationality rooted in plural approaches, and a scientific community with other skills, knowledge, and perspectives.

Acknowledgment

En commençant mon doctorat, je me suis plongé dans le chaos. Je ne parle pas du chantier hasardeux et des potentiels marasmes liés au doctorat, mais de la théorie des systèmes non-linéaires mieux connue du grand public sous le nom de théorie du chaos. J'y ai notamment appris que toute interaction déterminante au sein d'un système se retrouve encodée dans chacune des variables de ce système. J'ai d'abord trouvé cette pensée obscure, puis magique, voire folle. Mais à l'heure d'écrire ses remerciements, elle prend beaucoup de sens. Si cette thèse est en apparence le fruit de mes propres réflexions, elle intègre bon nombre d'interactions déterminantes que j'ai pu avoir par le passé, et ces remerciements n'ont d'autres buts que de dévoiler ceux et celles avec qui je les ai partagées, à savoir les co-auteurs indirects de ce manuscrit.

Je tiens d'abord à remercier mes parents, qui bien au-delà de simples conditions initiales, m'ont toujours encouragé à l'exercice de la pensée. Vous m'avez transmis le goût de la curiosité, celui du savoir. Vous m'avez appris à poser un jugement de l'esprit à travers un regard analytique mais aussi à travers celui du cœur et de ses valeurs, sans qui toute vérité perdrait de son sens. A mes frères et sœurs, ce cadeau vous l'avez également reçu et nous l'avons sublimé ensemble. Vous tous, vous m'avez toujours soutenu dans les épreuves et moments plus difficiles de mon parcours. Pour vous, sans oubliez mon grand-père, j'éprouve une grande admiration à l'égard vos oreilles et de votre cœur attentifs à mes théories et pensées loufoques, votre habilité à les comprendre, et à les commenter de façon à nourrir mes réflexions. Vous excellez bien plus dans ce rôle que les experts que j'ai pu rencontrer en conférence internationale.

Je n'oublie pas non plus mes ami.es, y compris mes collègues passés et présents que j'estime comme tels, ainsi que mes très chers cousins et cousines. Si j'ai pu avoir moins d'interactions quant à ma thèse avec certains d'entre vous, la théorie des systèmes et moi-même vous donnent une place toute aussi importante, car aucun système, en ce compris celui de la pensée, ne peut s'organiser sans relâcher et dissiper un excédent d'énergie, d'information, et de matière - excluons toutefois ce dernier - avec son environnement extérieur, en l'occurrence, vous. J'en conviens, c'est une bien médiocre définition de l'amitié, et elle ne traduit en rien toute l'estime que j'ai pour vous.

Avant d'entamer mon doctorat, j'ai pu apprendre de beaucoup et restituer ce que j'ai appris en me le réappropriant dans ce manuscrit. Je remercie mes professeurs du secondaire au Sacré-Cœur de Lindthout, en particulier, Mme Jans, Mme Bawin, et M. Yerlès, pour m'avoir transmis leurs passions respectives pour les mathématiques, la Grèce antique, et la philosophie. Je remercie également mes professeurs de l'UCLouvain qui ont fait de moi un bioingénieur, principalement, mes promoteurs de mémoire, les Pr. P. Defourny et M. Vanclooster, pour leur soutien et pour m'avoir respectivement introduits à l'analyse des systèmes et à l'hydrologie, ces sciences fascinantes. Le mémoire a été une épreuve difficile en tant que première expérience avec la complexité hydrologique et celle de la recherche scientifique. Cette occasion a néanmoins été celle de la rencontre de Jose Andres Ignacio avec qui j'ai pu collaborer dans le cadre sa thèse aux Philippines. C'était il y a plus de 5 ans que tu me remerciais dans ta thèse pour t'avoir montré comment surmonter les frustrations liées à la recherche. Aujour'hui, laisse moi corriger cette erreur sur la personne. Merci à toi, Andres, de m'avoir guider à surmonter les difficultés de mon mémoire. Cette énergie m'accompagne encore aujourd'hui. Tu n'étais pas le seul. Je remercie aussi Haykel Sellami, ton expertise sur l'analyse des incertitudes m'a été d'une grande utilité dans le cadre de ma thèse également. Enfin, je remercie ceux avec qui j'ai pu continuer à apprendre en tant qu'assistant de recherche pendant 4 ans au GERU et sans qui je ne me serais pas lancé dans une thèse, les Pr. M. Javaux et C. Bielders.

Je ne manquerai pas d'exprimer toute ma gratitude aux personnes direcetement impliquées dans ce projet de doctorat. Je remercie le FNRS, le personnel associé au Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA), ainsi que le jury FRIA, qui m'ont accordé cette incroyable opportunité de réfléchir pendant quatre ans et d'accéder au métier de chercheur. Pour cette riche expérience, j'exprime ma plus profonde reconnaissance à mes promoteurs, Marnik Vanclooster de l'UCLouvain et Michel Van Camp de l'Observatoire Royal de Belgique, qui m'ont offert énormément de liberté et toujours beaucoup soutennu, tant bien que mal, dans les trajectoires et attracteurs étranges de recherche que j'ai décidé d'emprunter. Je remercie les membres additionnels de mon comité d'accompagnement et de jury de thèse, le Pr. Hallet (UNamur), le Pr. Kaufmann (UMons), Pr. de Viron (LIENSs, Université de la Rochelle), le Pr. Crucifix (UCLouvain), et le Pr. Hanert (UCLouvain), pour les remarques et les regards interessés et enthousiastes qu'ils ont portés sur mon travail. Je

remercie les membres du projet KARAG pour le travail fourni en amont de cette thèse, en particulier les deux doctorants, maintenant docteurs, qui m'ont précédés, Arnaud Watlet et Amaël Poulain. Les chapitres de cette thèse ont également bénéficié de collaborations directes et très appréciées. A cet égard, je remercie le Pr. Feltz (UCLouvain) pour sa relecture et ses commentaires sur le chapitre 2, *Prof. Huffaker and Prof. Muñoz-Carpena for having cordially invited me to the University of Florida and for having trained me in nonlinear time-series analysis and sensitivity analysis*, Arnaud Watlet et Olivier Kaufmann pour leur expertise sur la tomographie de la résistivité électrique et leur collaboration aux chapitre 5 et 6 de la thèse, et le Pr. Olivier de Viron pour nos échanges, questionnements et collaborations répétées sur les méthodes d'inférence causale et pour l'invitation à les prolonger dans le cadre d'un projet de post-doctorat à la Rochelle.

Je terminerai en donnant de l'importance à toutes ces relations et influences qui s'ignorent. Je remercie ces hommes et ces femmes qui ont fait l'histoire des sciences, en particulier celle de l'hydrologie, et qui la portent encore aujourd'hui. J'ai cité quelques noms, mais je sais avoir manqué une myriade de personnes qui ont contribué et poursuivent aujourd'hui cette quête collective de science. Si vous vous y sentez investis, ces remerciements sont pour vous. En particulier, si vous lisez ces mots en tant que lecteur, si vous trouvez dans ce travail des réflexions qui vous accompagnerons dans vos démarches personnelles ou professionelles, vous donnerez plus de sens à mon travail que je n'ai pu lui en donné. Merci pour cela.

I will conclude by giving importance to all these relationships and influences that are unknown to each other. I would like to thank these men and women who have made the history of science, especially hydrology, and who still carry it today. I have mentioned a few names, but I know I have missed a myriad of people who have contributed to and pursuit this collective quest for science today. If you feet invested in it, these thanks are for you. In particular, if you read these words as a reader, if you find in this work reflections that will accompany you in your personal or professional endeavors, you will give more meaning to my work than I have been able to give it. Thank you for that.

Damien Delforge

x

Contents

List of Figures

List of Tables

List of Technical Abbreviations

AIC	Akaike Information Criterion (Eq. 5.5)
AMI	Adjusted Mutual Information (Eq. 5.7)
B&N	Brutsaert and Nieber Model (Eq. 1.5)
BIC	Bayesian Information Criterion (Eq. 5.6)
CCF	Cross-Correlation Function (section 6.2.1.1)
CCM	Convergent Cross-Mapping (section 6.2.1.2)
CMI(knn)	Conditional Mutual Information independence test (using a nearest-neighbor estimator) (section 6.2.1.3)
DAG	Directed Acyclic Graph (section 6.2.1)
EDM(-Simplex)	Empirical Dynamic Modelling (Simplex algorithm) (section 4.2.2)
ERT	Electrical Resistivity Tomography (section 3.3.2)
GC	Granger Causality (section 6.2.1.3)
GMM	Gaussian Mixture Model (section 5.2.2.3)
GSA	Global Sensitivity Analysis (section 4.2.3)
HAC	Hierarchical Agglomerative Clustering (section 5.2.2.2)
NSE	Nash and Sutcliffe Efficiency (Eq. 1.4)
ParCorr	Partial Correlation independence test (section 6.2.1.3)
(PC)MCI	Causal inference algorithm: (combining the Peter and Clark algorithm and) Momentary Conditional Independence testing (section 6.2.1.3)
PCA	Principal Component Analysis (section 3.3.2)
SI	Silhouette Index (Eq. 5.4)
TSC	Time-Series Clustering (section 5.2.1)

Other Acronyms

FNRS	Fond de la Recherche Scientifique (Belgium)
KARAG	Karst Aquifer Research By Geophysics (section 1.2.1)
LKS	Lhomme Karst System (section 3.1)
MIGRADAKH	MIning GRAvity DAta from a Karst Hydrosystem (section 1.2.1)
RCL	Rochefort Cave Laboratory (section 3.2)
ROB	Royal Observatory of Belgium
SPW	Public Service of Wallonia (Belgium)
UCLouvain	Université catholique de Louvain (Belgium)
UMons	Université de Mons (Belgium)
UNamur	Université de Namur (Belgium)

Chapter 1 Introduction

"All science is the search for unity in hidden likenesses [...] order must be discovered and, in a deep sense, it must be created. What we see, as we see it, is mere disorder."

Bronowski

Abstract

Hydrological systems can both be perceived and studied as very simple or very complex systems. This introductory chapter first covers how hydrological systems can be characterized and modeled in a simple way with lumped hydrological models or hydrograph analysis. Beyond that, they remain challenging to understand because of the numerous alternative modeling approaches that have been developed to deal with the specificity of hydrological systems worldwide, and the lack of observables to constraint and limit the number of models. Intrinsically, hydrological systems are complex because of their heterogeneity and a large number of interacting variables (high-dimensionality). Mechanistically, hydrological processes are sensitive to initial conditions (nonlinearity) due to this heterogeneity, their high-dimensionality, or the inherent nonlinear nature of hydrological processes, and the presence of thresholds ruling intermittent processes. If complex, they are nevertheless not random but organized, an emergent property of most systems allowing us to have both a generalized understanding of their behavior and to study them from various points of view that are not necessarily sophisticated. With empirical study cases developed on one of the most complex types of hydrological systems, i.e., karstic systems, the Ph.D. thesis investigates four topics: (1) What does it take to understand or know something, what is causality? (2) How to assess hydrological complexity from the nonlinear analysis of river recession hydrographs? (3) How to escape high-dimensions by identifying macro-level structures based on dynamic similarity? And (4) How to reveal robust causal relationships between hydrological variables?

1.1 Theoretical Background

1.1.1 Hydrological Systems: the Simple Way

Hydrology can be defined as a s*cience that deals with the waters above and below the land surfaces of the Earth; their occurrence, circulation, and distribution, both in time and space; their biological, chemical and physical properties; and their interaction with their environment, including their relation to living beings* (WMO, 2012). Since water involves necessarily interactions, hydrology is a systemic science. Systems themselves could be defined, among many other definitions, as an *assembly of objects, processes, or concepts, most often interacting with each other, which are focused on or result in a specific outcome* (WMO, 2012). The global hydrological cycle is a system. Still, systems are typically nested: they include many systemic units (e.g., terrestrial, oceanic, climatic, ecological, or social systems) while being themselves part of a broader system (e.g., the astrophysical system).

In terrestrial hydrology, catchments are typical systemic units closed by a watershed defined according to an outlet. Catchments are made of different interrelated components such as the surface, soils, groundwater bodies, the channel network, or human-made infrastructures (e.g., dams, pumping wells). Mostly focused on terrestrial hydrology, Dooge (1968) describes a system as *any structure, device, scheme, or procedure, real or abstract, that interrelates in a given time reference, an input, cause, or stimulus, of matter, energy, or information, and an output, effect, or response, of information, energy, or matter*. In contrast with the former, this definition stresses the importance of the input-output relationship to characterize a system. Most commonly, terrestrial hydrological systems are indeed characterized while considering the mass balance equation (or continuity equation in the form of an ordinary differential equation) which assumes that the state of an hydrological system is given by the variation of its water content over time (ΔWC, dWC/dt) responding to the total amount of water flowing in (Q_{in}) and out (Q_{out}):

$$\Delta WC \text{ (or } dWC/dt) = Q_{in} - Q_{out} \qquad Eq.1.1$$

This approach is not new. A first mass balance computation goes back to Perrault (1674) that succeed to close the annual mass balance of Seine basin in France ($\Delta WC \cong 0$) at the annual scale considering estimates of precipitation (Q_{in}) together with discharge and evapotranspiration (Q_{out}). He

concluded that rainfall was indirectly at the origin of springs. In doing so, he was at the same time reintroducing the modern vision of the hydrological cycle in Europe and pioneering modern hydrology (Bras, 1999). The mass balance is neither an outdated approach. For instance, the Gravity Recovery And Climate Experiment (GRACE) satellite mission allows to have a direct spatial and measurement of the variations of the water content (ΔWC), that is, a valuable closure to improve the global estimates of the mass balance components such as precipitation, evapotranspiration runoff or discharge (Wouters et al., 2014). Eq. 1.1 is extremely powerful to investigate any hydrological system and remains of an astonishing simplicity.

In the case of the hydrology of a catchment area, the change in water storage is usually the unknown, and the hydrological analysis of the catchment area usually involves the three types of data mentioned above: precipitation, evapotranspiration, and discharge. The simplest conceptual models consider at least two subsystems or conceptual reservoirs also constrained by the mass balance (Eq. 1.1): a surface system and a groundwater or aquifer system (Figure 1-1). The storage of the surface reservoir S_s is fed intermittently by the precipitation P. It loses water by evapotranspiration ET (surface evaporation and plant transpiration), through infiltration I (or recharge) feeding the groundwater reservoir, and through a quickflow Q_q that responds rapidly to rainfall. The Q_q includes rapid hydrological processes such as surface runoff (or overland flow), or rapid subsurface lateral drainage. Quickflow is intermittent and becomes negligible once rainfall has stopped for a while. The mass balance of this top reservoir is $\Delta S_S = P - ET - I - Q_q$. Since ET and P are input data, they can be grouped together into an effective precipitation input $P_{eff} = P - ET$. The groundwater reservoir storage only responds to $\Delta S_G = I - Q_b$, where Q_b represents the baseflow, or groundwater discharge that has a delayed response to rainfall events. The total river discharge Q is given by the sum of two components Q_q and Q_b.

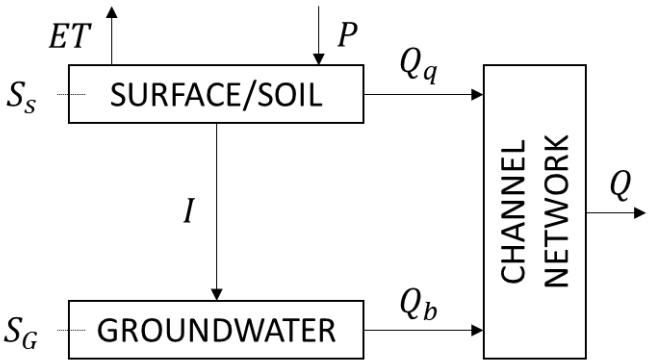

Figure 1-1: A simplified catchment hydrological system or lumped conceptual model. The model acknowledges three reservoirs: a Surface/Soil reservoir of storage S_S fed by precipitation P and emptied by evapotranspiration ET, recharge or infiltration I, and quick surface or subsurface flows Q_q; a Groundwater reservoir of storage S_G taking infiltration as input and releasing a baseflow Q_b; River discharge Q in the channel network sums up the fast contribution of quickflow and the delayed one of baseflow.

Figure 1-1 is a universal archetype, but more composite model structures are most of the time considered. For instance, one may take into account two separate reservoirs for the surface and the soil system. Then, the surface reservoir produces a quickflow assimilable to surface runoff, and the soil reservoir produces another flow representing the subsurface flow, usually referred to as interflow, with a delay in between quickflow and baseflow. Technically, it is possible to stack up reservoirs in a cascade to simulate the river discharge, as the Nash cascade model does with linear reservoirs (Nash, 1959). Intentionally, some would add them for their physical meaning: canopy water storage, different soils, or geological horizons. Whatsoever the number of modeling units, the mathematics of each abstract water reservoir is simple and aims at expressing the reservoir outflows Q_{out} (e.g., Q_q, I, or Q_b) as a function of the reservoir storage S. The relation is named the storage-discharge relationship. A common expression is:

$$Q_{out} = cS^d \qquad Eq.1.2$$

where c and d are parameters. If d is equal to 1, then Eq. 1.2 reflect a linear reservoir. Otherwise, the storage-discharge relationship is nonlinear.

Regarding, the groundwater discharge or baseflow, Eq. 1.2 is well acknowledged (e.g., Hall, 1968; Tallaksen, 1995). For the modeling of

quickflow Q_q, Eq. 1.2 should be applied considering a threshold storage value S_T to ensure the intermittency of surface or subsurface runoff processes:

$$Q_q = \begin{cases} c(S - S_T)^b, & \text{if } S > S_T \\ 0, & \text{if } S \leq S_T \end{cases} \qquad Eq.\,1.3$$

The excess water volume or depth, $S - S_T$, is the one that is routed relatively quickly to the channel network. The delay and the transfer time depend on the parameters c and d. It could represent runoff due to the exceedance of infiltration capacity (Horton, 1933) at the surface or due to the soil saturation and progressive rise of the water table (Dunne and Black, 1970). In catchment hydrology, Eq. 1.2 and 1.3 are not the most common equation to produce and transfer runoff amounts to the channel. Popular runoff production functions for catchment modeling are, for instance, the curve number method (SCS, 1972) or the Green-Ampt infiltration model (Green and Ampt, 1911). A typical routing function is the unit hydrograph (Sherman, 1932a), which behaves as a linear transfer function. In general, the underlying principles behind runoff activation are thresholds and exponential production responding to an effective precipitation followed by a delayed routing.

If one agrees on a conceptual model and its governing equations, then system identification is made when the optimal parameters are identified. One solution is to do it numerically with an implementation of the lumped catchment model and its equations. The hydrological model can be calibrated following a split-sample-test routine (Klemeš, 1986b) that divides the dataset into at least two periods. The first one is used for calibration, the second for validation. In most cases, the model is calibrated and validated by fitting the hydrograph, i.e., the total streamflow discharge Q (Figure 1-1). The fitting depends on an objective function. The Nash and Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970) is very popular:

$$NSE = 1 - \frac{\sum_{t=1}^{N}\left(Q_{sim}^{t} - Q_{obs}^{t}\right)^2}{\sum_{t=1}^{N}(Q_{obs}^{t} - \bar{Q}_{obs})^2} \qquad Eq.\,1.4$$

where Q_{sim}^{t} is the simulated hydrograph time-series of length N, Q_{obs}^{t} is the observed one with a mean $\bar{Q}_{obs}$. The highest NSE gives the best model, up to a limit of 1, indicating a perfect fit. A negative NSE indicates that the mean streamflow discharge is a better estimator than the model. If a good NSE is

both obtained for the calibration and the validation period, the model could be used for operational or scientific purposes. The parameters or the internal variables that were inferred (e.g., I, Q_q, Q_b in Figure 1-1) can be interpreted assuming that they are realistic. The model could also be used to design or test scenarios for scientific inquiry, water management, and support to decision making.

Another solution to hydrological system identification is hydrograph analysis. Hydrograph analysis is usually divided into two complementary procedures: recession analysis and hydrograph separation (Brodie and Hostetler, 2005; Hall, 1968; Tallaksen, 1995). Recession analysis is focused on the analysis of sustained decreasing streamflow periods, particularly low flows, such that the quickflow Q_q is considered null and the recession periods of the hydrograph Q data are reflecting the dynamic of the baseflow component Q_b. If no aquifer recharge (I in Figure 1-1) is further assumed during the same periods, one can estimate the parameters of Eq. 1.2 from by analyzing the dynamics of recession. Since the groundwater storage is usually not observed, recession analysis relies on other dynamical expressions of the recession. The Brutsaert and Nieber (1977) model (B&N) is the primary framework for recession analysis:

$$-\frac{dQ}{dt} = aQ^b \qquad Eq.\,1.5$$

where dQ/dt stands for the time derivative of the streamflow variable Q, theoretically supposed to reflect Q_b thanks to a recession extraction method. The recession constant a and the nonlinear exponent b are recession parameters. Since the aquifer reservoir is supposed to decay exponentially (Eq. 1.2), the same holds for the recession rate $-dQ/dt$. By applying a logarithmic transformation, Eq. 1.5 becomes linear:

$$\log\left(-\frac{dQ}{dt}\right) = \log(a) + b\log(Q) \qquad Eq.\,1.6$$

Conveniently, fitting a line in the recession plot cloud reporting the $\log(-dQ/dt)$ against the $\log(Q)$ allows estimating the recession parameters. Furthermore, a and b can be related to the aquifer reservoir characteristics, such as its thickness, porosity, or hydraulic conductivity (see Dewandel et al., 2003; Troch et al., 2013). Regarding the terminology, the term linear recession

refers to the case where $b = 1$, in which case the nonlinear exponent d of Eq. 1.2 is also equal to 1. However, the reservoir storage depletion or the recession dynamics remains exponential with respect through time. Indeed, the solution of the continuity equation (Eq. 1.1) for $dS/dt = -cS$ is given by:

$$\begin{aligned} S(t) &= S_0 e^{-ct} \\ Q(t) &= Q_0 e^{-ct} \end{aligned} \qquad Eq.\,1.7$$

Eq. 1.7 is widely used for linear recession analysis when a linear reservoir recession is assumed, and c is the recession constant ruling the decay $[T^{-1}]$ from an initial condition S_0 or Q_0. Similarly, the recession constant c can be estimated by applying the log transformation to Eq. 1.7 and fitting linearly individual recession segments.

The next step of hydrograph analysis is hydrograph separation. It consists of the decomposition of the hydrograph time-series Q into its two components Q_q and Q_b. Hydrograph separation could be done in situ using tracing tests or chemo-physical analysis. Ex-situ, hydrograph separation relies on recursive digital filters that incorporate the recession parameters obtained from recession analysis, eventually calibrated based on tracing tests (e.g., Chapman, 1999; Eckhardt, 2005; Stewart, 2015). Hydrograph separation allows estimating water budgets holistically and identifying the relative proportion of baseflow in streamflow (referred to as baseflow index), monitoring groundwater and managing it to ensure the reliability of water supplies to ecosystems and human activities, among other purposes (see Brodie and Hostetler, 2005). Moreover, hydrograph separation offers the opportunity to investigate quickflow related processes and parameters related to surface characteristics (e.g., Blume et al., 2007), and is often the preliminary step to model floods and rainfall-runoff (P_{eff} vs. Q_q) relationships (Beven, 2012b).

1.1.2 Hydrological Systems and Complexity

A simple conceptual model (Figure 1-1) formalized in a few relatively simple equations already provides a significant entry point to the study of the functioning of a hydrological system, whether through the reliance on lumped modeling, hydrograph analysis, or both (Kirchner, 2009). All these approaches are still widely used today. Nevertheless, terrestrial hydrological systems are often perceived or modeled as complex systems (see Bras, 2015;

McDonnell et al., 2007). Although widely used in science, complexity is not well defined, as noticed and discussed by Sivakumar (2017b). It is addressed here from four vantage points that help to explore the facets of complexity in hydrology.

1.1.2.1 Complexity and Understanding

A first intuitive definition of complexity would be “difficult to understand”, meaning “difficult to abstract or conceptualize”. It may also mean “difficult to predict”, however, a phenomenon may be easily predicted but hardly understood. To a certain extent, this definition is related to the person investigating the hydrological system. It may vary from person to person, just as what may be complex today for one person may become simple later on as long as the phenomenon is understood. It is, therefore, relative in time and from individual to individual.

For this reason, Sivakumar (2017b) does not consider such a definition as a workable definition. Objectively and scientifically addressed, complexity should be a trait related to the characteristics of the real system, that is not the abstract or conceptual one. However, the individual relativity of understanding remains a scientific issue. For instance, Refsgaard et al. (2006) show the significant differences in the outcomes when a hydrological systems tackle from different conceptual perspectives (Figure 1-2, see also Vanclooster et al., 2000). In this sense, the fact that several unequivocal models exist may be an argument for claiming that hydrology deals with complex systems. Indeed, when one deals with a simple physical system such as the oscillation of a pendulum, such a diversity of models does not exist. However, the plurality of models is problematic when hydrology is supposed to provide unbiased facts and pieces of information to societal stakeholders (Kirchner, 2017; Saltelli and Funtowicz, 2014). Concurrently, this problem is not specific to hydrology, and we can generally speak of science in a crisis of reproducibility, and consequently, in a crisis of credibility (Saltelli and Funtowicz, 2017).

In other aspects, the difficulty in understanding hydrological systems is not necessarily linked to the individual. It can be recognized at the community level, also for reasons that do not refer to the intrinsic properties of hydrological systems, but rather to the way we perceive or observe them. First, the relative complexity of hydrological systems could be imputed to the biased

perspective of scientific research. In that spirit, Beven et al. (2020) remind us that our perceptual understanding of hydrological systems (simplified in Figure 1-1) is quite good. However, we still expect that understanding in a science should grow over time (Sivapalan and Blöschl, 2017). This spirit drives scientists to engage in a complexity that Beven et al. (2020) call "sophistication" (idem for Sivakumar and Singh, 2012), but it usually results in improved predictions, and not in an overhaul of the basic perceptual model symbolizing our understanding. On the contrary, this ideal for improvement drives the development of a plethora of ever new and more sophisticated models that is sometimes referred to as cacophony (Dooge, 1978; Klemeš, 1982; Sivapalan, 2006), and which casts doubt on our fundamental understandings of hydrology (e.g., Figure 1-2), reinforcing this sense of dealing with a complex subject. It is therefore important to distinguish between the complexity of a system and that of a model, which is more reflective of the choices made by the modeler, whether explicitly stated or not, and to his/her attention to detail. Hence, what is complex, is perhaps hydrological systems, but also, in a subjective way, hydrological models themselves while diving into model sophistication.

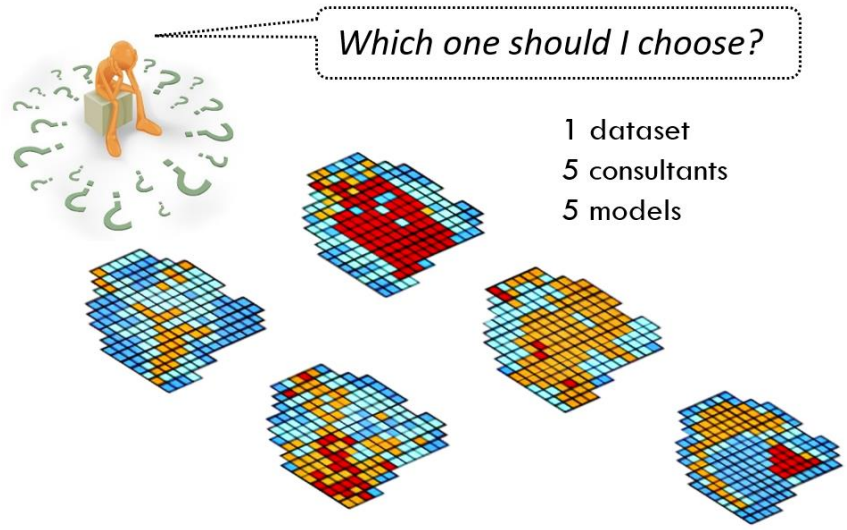

Figure 1-2: Individual relativism of modeling outputs (adapted from Refsgaard et al., 2006). The conclusions of the five consultants regarding groundwater vulnerability to nitrate pollution using five different models (red: very vulnerable; dark blue: well protected). As extrapolated scenarios, no data exist to validate the model predictions. However, models rely on the same input data, indicating sharp conceptual divergences.

In that regard, hydrological systems are also difficult to understand, because challenging to observe and monitor holistically. The subjective sentiment of complexity could be related to the lack of data and to the practical difficulty of observing the internal physical states of hydrological systems (Blöschl and Zehe, 2005), especially beneath our feet (Grant and Dietrich, 2017). The Prediction in Ungauged Basins (PUB) decade initiative was launched by Sivapalan (2003) to impulse the development of innovative strategies to assess the hydrological behavior of poorly gauged catchments (Blöschl, 2013; Hrachowitz et al., 2013). Currently, improving data and their acquisition technologies are seen as a crucial driver to make progress in hydrology (Beven, 2019a; Beven et al., 2020; Sivapalan and Blöschl, 2017).

In this respect, the ever-increasing availability of global hydrological datasets through satellite remote sensing offers new opportunities such as the direct monitoring of mass balance (Wouters et al., 2014), or the development of hyper-resolution global physically-based hydrological models (Sood and Smakhtin, 2015; Wood et al., 2011). The main challenges for these model of everywhere while diving into hyper-resolution is to remain relevant (Bierkens et al., 2015), given the remaining uncertainties in the remotely sensed dataset, the sustaining lack of knowledge about spatially distributed catchment processes and the importance of preferential flows in heterogeneous hydrological environments (Beven and Cloke, 2012). Indeed, it is now envisioned that physical representation of hydrological process for large catchment cannot systematically be derived from the laws derived from laboratory or hillslope experiments (Beven, 2000; Kirchner, 2006; McDonnell et al., 2007; Sivapalan, 2018). However, large scale hydrological experiments or testing is hardly conceivable as it is in general in Earth sciences. In particular, the physical representation of water flows in soils remains often based on Darcy-Richards equations that assume soil homogeneity. In some case, flows in an large heterogeneous medium can be modelled considering one homogeneous granular porous medium to account preferential flows, but their dynamics may deviate significantly from the descriptive behavior of this type of medium, making it impossible to obtain reliable results (Nimmo, 2009).

Preferential flows occurs due to local change in the hydraulic conductivity. This may occur due to soil heterogeneities funneling the flow into preferential flow paths, irregular soil moisture and air distribution patterns producing unstable flows, and very commonly due to the presence of macropores (ibid.).

These patterns of dual porosity are, however, ubiquitous, and explains the occurrence of fast flows and rapid spreads of pollutants where they should not appears (Beven and Germann, 2013). If at some point, it was expected that heterogeneities would disappear by pursuing finer spatial resolutions, Sivapalan (2018) reminds that they won't, whatsoever the scale. In particular, advancing the physical process representation of preferential flows remains in practice very difficult since the structural organization of heterogeneities and the macropore networks in hydrological systems are particularly hard to observe and characterize.

Finally, the difficulty of understanding hydrological systems could be due to their site-specificity, another reason for developing a plethora of models. All hydrological systems are unique simply because of their unique geographical location on Earth, an attribute called uniqueness of place (Beven, 2000). This uniqueness can extend to hydrological processes or the site-specific data. This fact has practical consequences related to hydrological modeling (see Beven, 2000). Uniqueness could be seen as an intrinsic property of hydrological systems. However, it is also a state of mind. From an epistemological point of view, considering that each hydrological system is a conceptually unique system is a major impediment to the development of a unified theory in hydrology. On the contrary, the conceptual uniqueness of hydrological systems would oblige hydrologists to develop specific approaches, and the growth of hydrology as a science would be an extensive collection of specific and empirical knowledge like in an encyclopedia. So far, the development of a new unified theory of hydrology is still on the agenda and actively researched (Clark et al., 2015; Kumar, 2007; Sivapalan, 2006), and pursued with calls for synthesis (Blöschl, 2006).

1.1.2.2 Complexity and Dimensionality

A system is made of elements in interrelation (Dooge, 1968). Then, if we look for an inherent property of systems characterizing its complexity, counting the number of variables that interact (that is, the dimension) is a walkthrough. Thus, if we consider the spatial and temporal variability of the meteorological and hydrological variables interacting with the heterogeneities of the landscape, subsoil, and geology over large areas, we can very quickly imagine a system of high or even huge dimensions. Conceptual models (e.g., Figure 1-1) can also quickly be accused of being unrealistic, excessively reductionist in the simplifying sense. At least since Freeze and Harlan (1969), many

hydrologists have moved away from the "system identification" approach in order to develop spatially distributed physically-based models pursuing more realism. These developments continue today but have been criticized. In practice, these models face the problem of equifinality of model parameters (Beven, 2006a). In other words, given the high dimensionality of the model and the lack of data to constrain it, many combinations allow good performance (e.g., Eq. 1-4). While predictions are often improved due to the model flexibility and its large degrees of freedom, the final parameters or internal state variables remain uncertain. Accordingly, Kirchner (2006) argues that such models are flexible marionette, correct in terms of hydrograph predictions but for the wrong reasons. In addition, physically-based spatially distributed models do not provide new understanding about catchment processes since they could be seen as a spatially distributed collection of lumped hydrological models (e.g., Figure 1-1; Figure 1-3) (Beven, 1989), based on the laws derived from laboratory and hillslopes experiments with no guarantees that they remain appropriate to describe processes occurring at large scale (Kirchner, 2006; McDonnell et al., 2007).

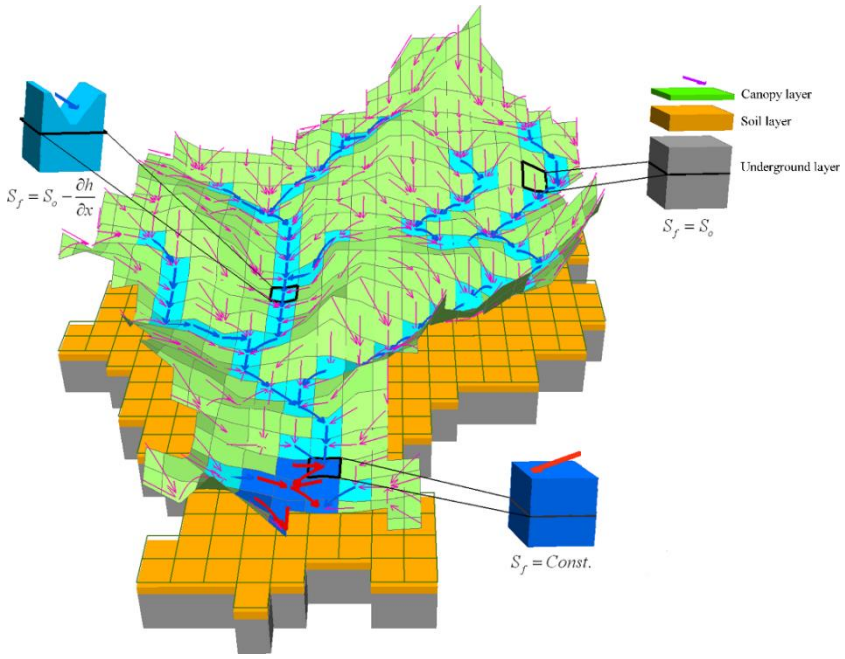

Figure 1-3: Example of a physically-based karstic catchment model for flood forecasting (from Li et al., 2019, CC BY 4.0). The physically-based model is spatially distributed and grid-based representation of connected lumped reservoirs models.

Grayson et al. (1992) further argue that the idea of a positive correlation between model complexity (i.e., sophistication, section 1.1.2.1) and confidence in the results is a misperception, in part due to a lack of discussion on its capabilities and limitations. Besides, they are usually far from parsimonious since the information in the model space (parameters and internal variables) is highly redundant. Hence, the dimension of a performant high-dimensional model overestimates by far the intrinsic dimension of the real system. However, the question remains as to how to assess the intrinsic dimension of a hydrological system. Naturally, it is conceivable that the dimension of a hydrological model must be matched with the intrinsic dimension of the real system to achieve an approach that is neither too simple nor too complex, but parsimonious. Regarding hydrological modeling, some new paradigms have emerged in that regard. The first way is to remove redundant information from the model (dimension reduction). For spatially distributed models, grid mesh showing similar surface, soil, or groundwater characteristics could be grouped into functional lumped units, e.g., hydrological response unit (Flügel, 1995). Applying clustering methods before or after model calibration is also an option (Pagliero et al., 2019). Using information-theoretic methods, Loritz et al. (2018) investigate model compressibility. They showed, first, that only a portion of topographic information in the model is relevant for the simulation of distributed runoff and storage dynamics. Secondly, compressibility is not time-invariant, and, arguably, dependent on the hydrological state of the system.

Besides, instead of picking predefined model, equations, or assumptions based on the modeler's choice (e.g., Figure 1-2), a model could be conceived based on an iterative approach that looks back and forth to the data, starts simply and progressively complexify up to a satisfying result (e.g., Fenicia et al., 2006). This flexible state-of-mind allows designing tailor-made or fit-for-purpose models while learning during the design instead of relying on high dimensional fit-for-all-purpose models that blur our understanding (Savenije, 2009). Nowadays, flexible hydrological frameworks allow testing rapidly various model architectures to select the optimal one (e.g., Clark et al., 2008, 2011; Fenicia et al., 2011; Gupta et al., 2012). Within such a model, the final number of internal variables could be interpreted as the dimension of the system, where each variable has a physical meaning.

Another approach would be to infer the dimension of the system from the output of the system, for instance, streamflow. First, signal decomposition

would allow to a certain extent by identify a number of component driving the signal. One can, for instance, identify harmonic components in the frequency domain of streamflow. This would allow us to identify the daily cycles in streamflow due to changes in solar radiation, temperature, and concurrently in evapotranspiration, as well as tidal or seasonal cycles. However, these cyclic components are ubiquitous in hydrological systems, and counting them will lead to the observation that hydrological systems have a similar dimension. They relate to patterns found in the external forcing rather than specific processes related to the hydrological functioning of the system. Considering these components as dimension would yield an underestimation. Indeed, they explain only a small part of the variability of the hydrograph, that in most case do not relate the hydrological processes of interest, which through spectral analysis, appears mainly as noise due to the random looking dynamic and occurrence of rainfall.

In particular, Sivakumar and Singh (2012) propose to address complexity from the scope of the theory of nonlinear dynamic systems, or chaos theory (see Sivakumar, 2017a). An essential part of the chaos theory relies on the systemic concept of state space reconstruction (Deyle and Sugihara, 2011; Packard et al., 1980; Sauer et al., 1991; Takens, 1981). The state space is the D-dimensional space whose coordinate axes are represented by the D explanatory variables of the system, where D would be the physical dimension of the system. Within the state space, all states and trajectories (or orbits) visited by the system can be reported. The problem is that, generally, one cannot study the actual state space since the internal variable of the system are usually unknown (e.g., Figure 1-1). However, Takens, (1981) provides a popular method for state space reconstruction from only one variable time-series, referred to as the delay embedding theorem.

Intuitively, in an abstract and deterministic dynamical system ruled by a few ordinary differential equations, the state of the system will define the future trajectory of each variable (e.g., Eq. 1.1 to 1.3). Hence, the trajectory of one single variable is related to the state of all variables in the system. Takens thus proposes to analyze the trajectories of one single variable, recovered using successive delays, to reveal the dynamical structure of the whole system (Figure 1-4). The necessary number of delayed coordinates to unfold the dynamic is called the embedding dimension (m), which could be interpreted as a complexity metric, however, potentially different than D (see section 4.2.2.1). Sivakumar and Singh (2012) propose instead to estimate the fractal

(i.e., not integer) dimension of the reconstructed systems, which reflect how the attractor manifold (the geometry drawn by the trajectories) tends to occupy the reconstructed state space. In the case of the Lorenz's system (Figure 1-4), the fractal dimension is slightly above 2 because the wings of the butterfly extend themselves as a surface. In fine, Sivakumar and Singh (2012) propose the fractal dimension as a generic estimator of dynamic complexity that could guide the modelers in building models with a dimension that correlates with the intrinsic dimension of the modelled dynamic.

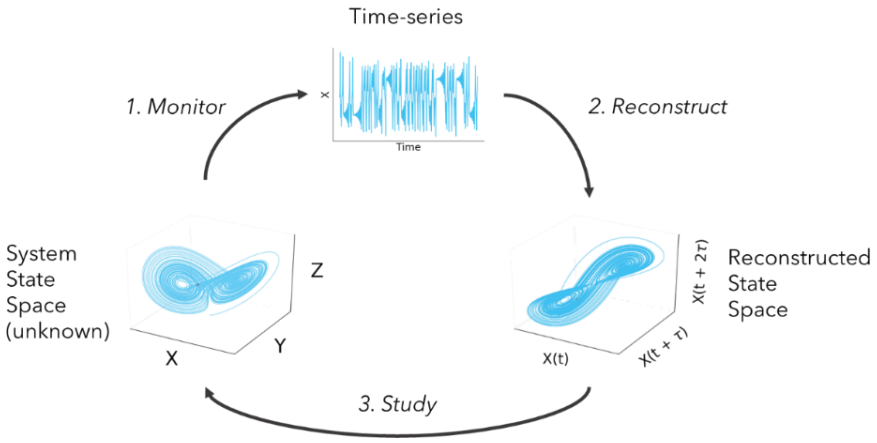

Figure 1-4: Reconstruction of system dynamic from one time-series with Takens embedding theorem. The example shows the emblematic Lorenz system (1963), a simplified model of atmospheric or thermal convection, whose trajectories depict a butterfly. It is possible to reconstruct a pseudo-system with the same characteristics from a single time variable and its successive lags (Takens, 1981).

Finally, dimensionality, even if it is a characteristic that can be attributed to the real system, is related to complexity in a sense "difficult to understand" (section 1.1.2.1). Quite directly, a large number of variables impede the generalized conceptual understanding of the system. According to Koutsoyiannis (2006b), the quest for small dimensions of hydrological dynamics nourishes the hope of representing a complicated system with a few equations. The author nevertheless recalls the limitations of the theory of nonlinear dynamic systems and the necessity to discuss them with respect to the characteristics of hydrological time-series such as intermittency, noise, the time-series length, or the effect of high auto-correlation (also reviewed in Sivakumar, 2017a).

1.1.2.3 Complexity and Mechanisms

Within a system, mechanical processes can reduce or expand the dimension of the dynamic if compared between the input and the output. The fact that streamflow data could be modeled with a simple conceptual model, incomplete data, and few ordinary differential equations (section 1.1.1, see also Kirchner, 2009) is proof by example. Indeed, rainfall is undoubtedly much more a random-looking both in space and time, high dimensional and less predictable than streamflow. A simple way to understand that is to look at Figure 1-5. The first device (a) is a Quincunx as designed by Galton (1889). Quincunx refers to the structure of offset obstacles through which marbles must pass to end up in a compartmentalized receptacle. Thanks to a funnel, the balls start their passage under similar initial conditions. Nevertheless, they remain sufficiently different to generate at the end, a random distribution that follows a normal law. The other device (b) is a conceptual rework such that the device is able to accept any random input and turn it into a deterministic outcome. These are drastic views. Most likely, a hydrological system such as a catchment exhibits both characteristics. Similar rainfall events occurring during similar environmental conditions will still produce some variability in the hydrograph response due to sensitivity to initial condition or high dimensional interactions. Conversely, similar hydrograph response may occur with different rainfall event at different environmental conditions. After all, a hydrological catchment or any closed drainage system defined by an outlet is like a funnel.

Generally speaking, sensitive dependence to initial conditions (e.g., Figure 1-5.a) is referred to as nonlinearity (Sivakumar, 2017b). All nonlinear dynamical and deterministic systems exhibit such a dependence up to the point they may yield to random-looking outputs. This is chaotic determinism. Even simple systems, such as the Lorenz system with three variables (Figure 1-4), could be chaotic. Hence, randomness is not solely imputable to high-dimensionality (section 1.1.2.2). Consequently, the presence of nonlinearity affects the predictability of a system over the long term.

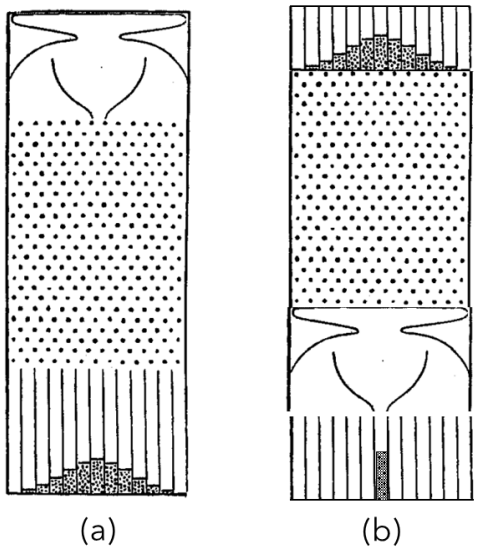

Figure 1-5: The mechanics of randomness and determinism. (a) the original Galton Quincunx (1889) able to generate a Gaussian random distribution from tiny differences in the initial input; (b) a reversed conceptual device able to generate a deterministic output from random input.

Hydrological systems are inherently nonlinear as being sensitive to initial conditions (Sivakumar, 2017b). However, one would always expect some linear correlation between catchment input (effective precipitation) and its output (river discharge). Linearity implies strictly proportional relationships. Hydrological systems are, therefore, nonlinear but still present mainly monotonic relationships. For this reason, hydrology has mostly relied on linear methods to understand hydrological systems (e.g., Dooge, 1973), especially before the 2000s (Sivakumar, 2017b). According to Blöschl and Zehe (2005), the use of linear methods can explain the occurrence of poor predictions in hydrology. Otherwise, the nonlinear nature of the system itself affects its predictability, together with the lack of observability (section 1.1.2.1). Nonlinearity in hydrology takes forms that have already been mentioned. First, nonlinearity is mainly associated with curvilinear or exponential relationships (e.g., Eq. 1.2, 1.5, 1.7, Brutsaert and Nieber, 1977; SCS, 1972). Typically, the groundwater hydraulic theory of groundwater suggest that the discharge of homogeneous aquifer is nonlinear (Eq. 1.5 with b=1.5), but higher degrees of nonlinearities are found when the aquifer geomorphology is heterogeneous or anisotropic (Brutsaert and Nieber, 1977; Harman and Sivapalan, 2009b; Rupp and Selker, 2005; Troch et al., 2013).

Nonlinear recession patterns (Eq. 1.5) may result from the joint contribution of different reservoirs. For instance, the linear discharge (Eq. 1.7) of heterogeneous hillslopes in the landscape may explain the nonlinearity of the total streamflow in large catchments (Chen and Krajewski, 2015; Clark et al., 2009; Harman et al., 2009).

The other type of nonlinear mechanism are thresholds, which are crisp sensitivity to initial conditions behind the intermittency of hydrological processes (e.g., Eq. 1-3). Thresholds are behind processes such as surface runoff or macropore flow activation (Blöschl and Zehe, 2005; Wilson et al., 2017; Zehe and Blöschl, 2004; Zehe and Sivapalan, 2009). In more general terms, they are conceptually behind any processes that may occur at some level of saturation or desaturation of a reservoir, true or abstract, affecting an hydrological system (e.g., relative humidity and condensation or rainfall, dam's spillway system, legal or illegal emergency reactions during floods or droughts). Thresholds or intermittency are problematic in terms of nonlinear time-series analysis. An intermittent process repeatedly switches from a dimension of 0 (a constant point, no trajectory) to a higher dimensional and nonlinear dynamic and vice versa. For this reason, intermittency leads to the underestimation of the dimension of intermittent dynamic (Koutsoyiannis, 2006b; Sivakumar, 2001). Intermittency further violates the hypothesis that nonlinear dynamical systems never revisit the exact same point in the state space, i.e., the hypothesis of aperiodicity (Kantz and Schreiber, 2003; Sivakumar, 2017a). Surely, if one reconstructs a system (Figure 1-4) from an intermittent variable, he/she will not have information about the system states while the process is off.

Nevertheless, threshold effects are sensitive to resolution and scales both in time and space. A daily rainfall time-series may contain approximatively 50% of zero values in Belgium, none in a monthly series. Spatially, thresholds are supposed to be less crisp over a large area because different thresholds occur at different places and times. Indeed, they could even disappear if we consider that a large watershed with perennial stream encompasses some of its subbasins that have intermittent streams or runoff axes. In that spirit, Brutsaert and Nieber (1977) already showed that the early stage of recession exhibit higher nonlinearity, most likely related to the presence of intermittent flow. Similarly, but from a spatial point of view, nonlinearity evolves within a catchment with changes in the extent of the open channel drainage network (Biswal and Marani, 2010).

As with dimensionality, the nonlinearity of natural processes can affect the processes of understanding (section 1.1.2.1). In simple terms, understanding favors linear explanations, even semantically, such as “rain causes floods”, or “wind increases evapotranspiration”. Whether someone who would respond to this assertion by saying “it depends”, although the answer is more correct than the assertion itself, the answer is far from a satisfactory explanation and most will agree on a “yes”. Linearity is interpretable by the brain and easily communicated. The very definition of nonlinearity is, in this sense, linear: the more sensitive dependence, the more nonlinearity. The popularity of the B&N model (Eq. 1.5) is also because, beyond its simplicity (section 1.1.2.2), it is linearizable (Eq. 1.6). Pragmatically, a scientist that uses nonparametric nonlinear methods may expect better predictability but will have more trouble to interpret and communicate the results, especially if the method is nonparametric, because it usually means that the model structure is not specified into a comprehensive and interpretable equation (section 1.1.2.1), and high dimensional (section 1.1.2.2) as many methods developed in the field of machine learning (Shen, 2018).

1.1.2.4 Complexity and Organization

Systems are typically organized, and another way to address complexity is to refer to the complexity of such an organization. Organization has various and often conflicting definitions (see Sivakumar, 2017b). In general, organization is a feature attached to the whole system. It presupposes some kind of order that arises from the general properties of systems: (1) *boundaries* or constraints, (2) while remaining *open* to allow storages and exchanges matters, energy, or information from and towards their external environment, and (3) internal interactions including *feedback* processes. Since the system contributes to its own level of organization, we speak of self-organization. From a physical point of view, dissipative systems are open systems that share these properties. Nowadays, catchments are increasingly seen as dissipative systems (Kumar, 2007). They have boundaries in space (e.g., the watershed), but also in time since hydrological time-series are bounded by their extremes. They exchange water bi-directionally with the atmosphere, store water, release it downstream, and involve processes that include feedbacks and interactions (Figure 1-1).

In practice, and often without resorting to the concept, organization is described in various ways. A general word that is often used in hydrology and

relates to organization is the term pattern (Sivapalan, 2006). To assess organization or patterns, it is first possible to describe static spatial features, such as the geometrical properties of the watershed or the surface drainage network, the spatial arrangement of topography, land cover and use, pedology, geology in an extensive qualitative way or relying upon summarized statistics or indicators. Similarly, one may look for time-invariant properties in the hydrological time-series, for instance, the statistical distribution or its corresponding moment. In particular, the cumulative probability of streamflow, also known as the flow duration curve, is an important hydrological signature characterizing a catchment (see, for instance, Figure 3-3). Computing the long term relative contribution of flows to the mass balance or the hydrograph is also an option, e.g., done by the baseflow index revealing the long term importance of groundwater discharge in the streamflow ($\bar{Q}_b/\bar{Q}$). One could also estimate chaotic invariants such as indicators of the fractal dimension, or the sensitive dependence to initial conditions. In a way, the dimension reduction that is operated between the input (effective precipitation) and the output indicates a gain of order and thus reflects the organization of the system. Another way to picture organization statically is to present the causal relationship in a conceptual model (e.g., Figure 1-1). In general, when the causal relationships in a system are reported with directed arrows between variables, we speak of directed acyclic graphs or DAG (Pearl, 2000).

Organization is also dynamic and varying on the short terms (e.g., Loritz et al., 2018). Then, organization is also reflected by the variations found in the time-series of the different hydrological components. They are revealed, temporally, spatially or both, by relying either on modeling, signal decomposition, or hydrograph separation. Historically, hydrology has been focused on the study and the reproduction of temporal patterns, e.g., by fitting the hydrograph. Accordingly, Sivapalan (2018) refers to it as the paradigm of hydrograph fitting. However, now that spatial data is becoming more and more abundant, hydrologists are more interested in studying the spatiotemporal patterns as well (Woods, 2002). In particular, the concept of hydrological connectivity (Bracken et al., 2013) highlights that not the entire area covered by the watershed contributes to discharge, but only the areas connected to the outlet, dynamically, depending on the hydrological state of the basin. In this spirit, the watershed could be an improper boundary for the system closure, which could be time-variant.

Finally, organization transcends scales as systems are typically nested into each other. Understanding the up or downscaling behavior of variables and processes both in space and time is an important topic in hydrology (Blöschl and Sivapalan, 1995; Klemeš, 1983), that has already been partially introduced with respect to the scaling behavior of dimensionality, a nonlinear processes or intermittency (section 1.1.2.2 and 1.1.2.3, and above).

As with dimensionality or nonlinearity, organization, while an inherent property of natural systems, also affects understanding. First, for the reasons that were already discussed, because organization encompasses interrelations that are mechanically described (1.1.2.3) and involve several variables (1.1.2.2). Hence, a high-dimensional and nonlinear real system will be difficult to abstract into an organized conceptual one symbolizing our understanding. The other reason would be that understanding could also be seen as a system that includes boundaries, constraints, inputs, outputs, storage, exchange, interrelations, and feedback. In that sense, this Ph.D. is an output of a system, much larger than myself, that targets understanding, and I hope that the lower level boundary elements introduced in this background will allow understanding this work and its purpose, hopefully, with some level of organization.

1.2 Ph.D. Project and Dissertation

1.2.1 The MIGRADAKH project

The Ph.D. project started in January 2017 funded by a FRIA grant provided by the Fonds de la Recherche Scientifique (FNRS, Belgium). The project was launched under the name “MIning GRavity DAta from a Karst Hydrosystem” (MIGRADAKH). The Ph.D. project follows another project started in 2013, Karst Aquifer ReseArch by Geophysics (KARAG, www.karag.be). KARAG allowed the hydro-geophysical investigation of the site studied in this Ph.D., which is the Lhomme Karst System (LKS, Chapter 3) next to the city of Rochefort (Belgium). KARAG has resulted in the publication of two Ph.D. dissertations firming up the understanding of the hydrological behavior of the system (Poulain, 2017; Watlet, 2017).

The MIGRADAKH project was motivated by the fact that hydrology has always been mainly relying on a hypothetico-deductive approach. Whether based on simple (Figure 1-1) or more complex physical and possibly spatially

distributed models (e.g., Figure 1-3), these approaches assume how the system works, while its specificity is obtained through parametrization. Today, hydrology is turning more towards empirical approaches (Sivapalan, 2018) to the extent that our conceptual understanding is relatively uncertain beyond the basics (section 1.1.2, Figure 1-2).

Initially, the goal of the Ph.D. project was to investigate new empirical, dynamic, and causal inference methods from the theory of nonlinear dynamical system to assess how empirical causality detection from time-series could recover hydrological or other causal connections in the LKS. The second objective was to investigate further how these empirical techniques could be used to develop physically-based karst models with a flexible data-driven structure and closing the mass balance using local gravimeters (Delforge et al., 2017). However, the idea of being able to solve causality using an empirical method and nesting it in a physically realistic model was a bit presumptuous. Indeed, somehow, unveiling the causality of the hydrological cycle is the ultimate goal of hydrology (Blöschl et al., 2019). In investigating these methods, the Ph.D. ran up against their limitations, and the project remained focused on the potential of these methods.

1.2.2 Thesis Outline

1.2.2.1 Scope: Karst Systems and Beyond

Among hydrological systems, karst systems are perceived as very complex ones. Their studies are particularly relevant in that they provide about 25% of drinking water supplies, occupy around 10% of the world's continental surface area (Ford and Williams, 2007), and are present in 21.6 % of European geology (Chen et al., 2017). Their complexity comes from the fact that karst systems develop on soluble rocks and, as a result, a particularly heterogeneous network of micropores and macropores (cracks, fissures, conduits, or caves) develops below the ground (Bakalowicz, 2005; Goldscheider and Drew, 2007; Hartmann et al., 2014; White, 2006). Beyond heterogeneity, karst systems exhibit the following features (ibid.; see also karst.iah.org):

1. Evolution: they change over time and possibly with abrupt changes, for instance, through the collapsing, dissolution, erosion, or the ghost-rock karstification (Dubois et al., 2014) processes;

2. Individuality (i.e., Uniqueness): because of the singular heterogeneities and the network organization of each karst system, their structure and dynamics are particularly unique (i.e., less generalizable);
3. Lack of observability: the heterogeneities cannot be exhaustively characterized, and karst studies have to deal with insufficient databases;
4. Anisotropy: hydraulic properties vary depending on the orientation of the geological materials and the fracture network;
5. Duality of porosity: karst systems present at least a duality of porosity considering matrix microporosity and the macropore network. The duality of porosity also leads to the duality of flow (e.g., quick preferential flow and delayed matrix baseflow).
6. Nonlinearity: the sensitive dependencies to initial conditions result from the heterogeneities, a higher number of natural reservoirs (high dimensionality), or duality of porosity and their threshold fill-and-spill effects.

So far, all these characteristics, although exacerbated in the case of karst systems, are now seen as universal and challenging characteristics of hydrological systems (section 1.1.2). Furthermore, Hartmann et al. (2014) report that karst systems are still studied and modeled from the same simple lumped basis as exposed in section 1.1.1, but also from a wide variety of model (section 1.1.2.1, Figure 1-2) ranging between these simple lumped conceptions to high-dimensional spatially distributed, and physically-based (section 1.1.2.2) representations, with similar scaling issues (section 1.1.2.4, see also Kiraly, 2003). Finally, the concept of the watershed as a closure of the hydrological system is also challenged (section 1.1.2.4) in the case of karst systems (e.g., Bakalowicz, 2005). These are regularly delineated through the concept of connectivity, for example, by means of tracer tests (ibid.).

Then, two reasons allow explaining why the scope of this thesis expands itself beyond karst systems. The first one is that, undoubtedly, from any initiative focusing on the conceptualization of karst systems will derive useful applications to understand hydrological systems in general. Secondly, the Ph.D. project is about conceptualization using empirical methods applied to time-series in general, that can be applied to all hydrological systems if not all systems monitored over time.

1.2.2.2 Thesis Content and Organization

With empirical methods, the thesis explores, although non exhaustively, the four domains discussed in section 1.1.2, that are understanding, dimensionality, mechanics, and organization. As a second introductory section, Chapter 2 mines further the problems related to the development of a robust understanding of hydrological systems. For that purpose, the chapter is an epistemological essay that attempts to define what is causality based on the historical evolution of the concept in the philosophy of sciences, however, illustrated from the scope of hydrology. In Chapter 3, the study site and the data used for discussing the other themes are presented. In particular, Chapter 4 develops a method to address catchment complexity by analyzing the nonlinear patterns in river recession data in the Lhomme catchment. Chapter 5 is a comparative analysis relying on Time-Series Clustering (TSC), i.e., dimension reduction, to identify the major hydrological and lithological structures (hydrofacies) in the subsurface above the Rochefort caves from a high-dimensional electrical resistivity model. The last chapter investigates causal inference methods and their ability to reveal system organization as a causal graph. It combines a virtual experiment, a case studying hydrological connectivity between the surface and the percolation in the underlying cave, and a case focused on the general behavior of the Rochefort cave system relying on relative gravimetry data monitoring mass changes in the system. Since I study the surface river network, the karst subsurface, and the percolation patterns successively, the thesis structure can also be interpreted as diving progressively more in-depth into the karst.

Chapters 2, 4, 5, and 6 are adaptations of scientific papers that were or will be submitted in the frame of this Ph.D. project. Publishing in a scientific journal most of the time, requires a specific and narrow focus on a specific problem, the scope of the journal, and the targeted audience. The particularity of these sections shall not hinder the understanding and follow-up of the general message of this thesis. To ensure it, these chapters are all preceded by a Foreword section allowing me to depict the broader context of these adapted publications. Moreover, such informal sections provide a storytelling angle to the thesis and report with transparency the initial intentions behind the research. From an epistemological point of view, the next chapter will cover why such transparency is also essential in science.

Chapter 2 Which causality for hydrology? An evolutionary perspective

"When we try to pick out anything by itself, we find it hitched to everything else in the Universe."

John Muir

Foreword

Initially, the general and naive idea of the thesis was to use empirical methods to reveal the causal links, their arrangement, to build a physical model articulating on the revealed causal structure (section 1.2.1). The method of causal inference was the Convergent Cross-Mapping (CCM) method, which takes into account the nonlinear interactions between variables and, therefore, interesting to apply on a karst system. The first tests of this method revealed causal links that were difficult to support, such as local hydrological variables affecting precipitation. These illogicalities pushed me to feed an interest in the definition of causality, aside from my scientific investigations.

Ideally, any starting point for a method is a definition that states principles. Then it is easier to think about how to turn them into a methodical approach. In scientific papers, the definition of causality is often avoided or barely introduced. Causality is succinctly linked to the identification of a cause and its effect, their mechanism, or defined in the negative, e.g., correlation is not causality. I believed that causality is closely related to epistemology, which studies knowledge and its limits. So, I conducted a parallel review of causality in the philosophy of science and a review of epistemological concerns in hydrology. Over the first two years of the project, my view of causality, and the way I would have defined it, changed constantly. It began to stabilize with the progressive writing of this chapter, started in summer 2019, followed by many exchanges with my supervisors. After valuable remarks by Bernard Feltz, philosopher of science at UCLouvain, the writing ended a year later. He helped me to sharpen the conclusion and chase away the few inconsistencies. A simplified version of this chapter is being considered for publication in the opinion section of the HESS journal.

Abstract

Causality is a vague and controversial concept but persistent in the sciences. There is currently a growing interest in it. For the science of hydrology seeking to unify its theory of hydrological systems, this opinion assumes that a unified understanding of causality is the first requirement. Through a Darwinian evolution of causality in the Philosophy of Science from ancient Greece to the present day, causality is related to contemporary epistemological topics in hydrology such as the debate between theoretical or applied hydrology, the problem of change, the realism of models, whether physical or empirical, the different causal perceptions of hydrology among hydrologists and within society, or the view of hydrological systems as self-organized structures with emergent functions. Through the journey, the most recent and robust methods for measuring causality are reviewed, depending on whether causality is related to physical rationalism, empirical observation, psychological or societal perceptions of facts, or to an organized and systemic behavior. While all these approaches have fruitfully contributed to advances in hydrology and to the way causality is thought, there is a need for a definition of causality that encompasses them. As suggested, causality is grounded in the minds of a community that has integrated the past interactions with the real world into facts, theories, and methods characterizing its ability to infer further knowledge in the future and operate in the real world. Causal explanations emerge from this evolving system as stable, robust, constrained by logic, and testable consensus on how systems are conceptualized within the scope of science that fixes its context, limits, and purposes. More than a body of causal laws or theories, we argue that causality today has become a flexible analytical framework to guide and frame science in its quest for understanding and progress.

2.1 Introduction

Which causality for hydrology? This idea of a plural causality contrasts with a conventional interpretation that expresses one intangible truth about the mechanisms of a phenomenon, and for which science should work to unveil their nature. Still, causality is a vexed question, and this definition is one of many. It has plural meanings inherited from its controversial history. Formalized in Ancient Greece, Aristotle suggested four types of causes aiming to provide the best explanations (Falcon, 2019). In the 18th century, Leibniz formalized the principle of sufficient reason, which states that any phenomenon in the real world shall have natural causes (Melamed and Lin, 2018). At this time, causality was instead interpreted temporarily by the directed link between two successive events, the cause and its effect. However, the philosopher David Hume has shown that such a link cannot be demonstrated, relegating causality as a law to the rank of belief (Hume, 1738). In 1912, the mathematician, logician, and philosopher Bertrand Russell pointed out that the use of the term causality or cause, at that time, had disappeared from science and was just discussed among philosophers: "*the law of causality [...] is a relic of the bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm*".

Today, three centuries after Hume and one after Russell, what is the status of causality in science? Figure 2-1 reports the trends in the percentage of scientific articles that use the specific terms "causality", "causal", "causative", and "causation" in the title, keywords, and abstracts of scientific articles within the SCOPUS literature database (Appendix I for further details).

All linear trends are significantly positive (p-value < 10^{-2}), suggesting a come-back either in the scientific interest or affinity with the concept of causality. Regardless of the domain (ALL), scientific articles refer today twice more than 20 years ago to the causality related terms in their titles, keywords, and abstracts. The relative use of the words is much more pronounced in Social Sciences (SS), Heath Sciences (HS), and Life Sciences (LS) articles compared to Physical Sciences (PS), or specifically to hydrological or water-related journals (HYDRO) given the historical influence of physics and engineering in hydrology. Several hypotheses can explain this schism. Russel's opinion on a science free from the vagueness of causality, thanks to its method, may still be present in the Physical Sciences. Besides, speaking of a physical causal

relationship is somehow tautological. One could argue that in PS, physical means causal since a non-causal physical relationship is a non-sense. Paradoxically, the reliance on a causality-related terminology is the strongest in what are sometimes called inexact sciences (SS). Preferably, it could relate to system complexity: the more complex systems a discipline studies, the more causal terminology it uses. This potential pattern would explain the percentage differences between the HYDRO group and the AGRI and ENVI groups, both including publications related to the science of hydrology, but more focused on systemic concerns that involve more complex interactions.

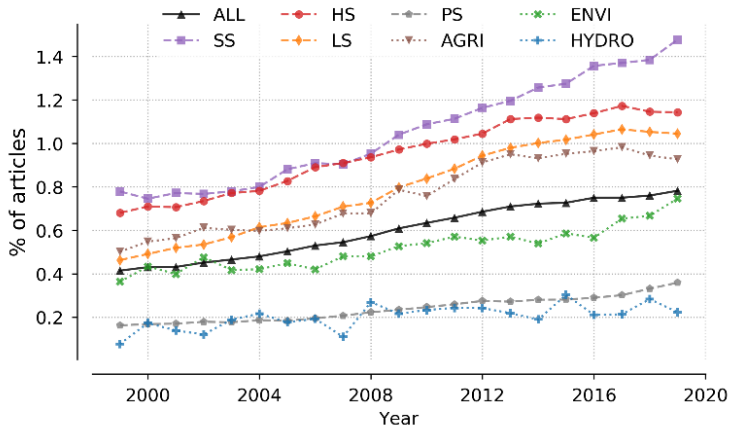

Figure 2-1: Trends in causality related terms in the titles, keywords, and abstracts of scientific articles (1999-2019) for all SCOPUS articles (ALL); the scientific domains of Social Sciences (SS), Health Sciences (HS), life sciences (LS), and physical sciences (PS); the specific subject areas of Agricultural and Biological Sciences (AGRI ∈ LS) and Environmental Sciences (ENVI ∈ PS); for articles whose source title contains the prefix 'hydro' or the word 'water' (HYDRO).

In this opinion, causality is hypothesized as the cornerstone of all science, but its facets point to different conceptual meanings, making science seemingly fragmented. What is tacitly or explicitly meant by causality in the diverse field of science or between individual scientists is possibly different.

The same assumption could hold for hydrology, which is often described as fragmented science:

1. Into sub-disciplines, based on the component of the hydrological cycle (e.g., surface, subsurface, groundwater hydrology). According to Klemeš (1982), the purpose of fragmenting sciences into disciplines and sub-disciplines is to split the possibly infinite causal chain into segments, each containing only a few chain links;
2. Into a physical fragmentation due to the uniqueness of every hydrological environment on Earth, further leading to their uniqueness of processes (Beven, 2000). This view considers that every particular system has its own causal laws threatening a unified understanding of hydrological systems.
3. Between operational hydrology oriented towards engineering, management, decision making, or forecasting, independently of complete understanding of hydrological processes, and the science of hydrology which aims to understand these processes and expect to become autonomous regarding its hydraulic engineering background (e.g., Clifford, 2002; Dooge, 1968; Klemeš, 1986b). The dichotomy depicts two kinds of causality: one oriented towards pragmatism and the other towards understanding;
4. Into a cacophony depicted by the jungle made of many physically-based hydrological models (Dooge, 1978; reiterated by others: Klemeš, 1982; Sivapalan, 2006) that have been developed since the physically-based paradigm introduced by Freeze and Harlan (1969). The latter is motivated by the idea that causality should specifically relate to physics to remain useful in extrapolated environmental conditions. Yet, dozens of physically-based models exist (see Beven, 2012; Singh and Frevert, 2002a, 2002b), from simple lumped ones to hyper-resolution spatially distributed ones, all-carrying causal views on their own. Complex models were supposed to be more physical as they tend to explain the hydrology from detailed discretized parts of the system and could, therefore, be attached to a reductionist, mechanistic approach referred to as Newtonian, bottom-up, or upwards approaches to causality (Klemeš, 1983);

5. Between empirical approaches, often more related to operational hydrology. Such methods have more emphasis on data modeling rather than physical process representation: data-based modeling (Young, 2013), stochastic modeling (Koutsoyiannis, 2016; Yevjevich, 1987), chaotic deterministic approaches (Sivakumar, 2017a), information-theoretic approaches (Goodwell et al., 2020; Singh, 2015), or deep learning algorithms (Shen, 2018). If Klemeš (1982) criticized these "let-the-data-speak" approaches, they have never disappeared and may be carried out by researchers that have faith in their contribution to the causal understanding of hydrological systems.
6. Between people: first, scientists since the methods mentioned above and scope of investigation divide researchers and possibly their opinions. Modelers using different models will come up with very different outputs and conclusions (Refsgaard et al., 2006; Vanclooster et al., 2000; Figure 1-2). This fact is concomitant with a general reproducibility and, therefore, a credibility crisis in science (Saltelli and Funtowicz, 2017), and a further fragmentation between science and society. Some people do not trust scientists anymore in their neutral and objective ability to reveal facts and their causes (Kirchner, 2017).

A genuine opinion about the fragmented status of hydrology should be balanced considering the numerous initiatives that seek to unify hydrologists around a theory (Beven, 2006b; Dooge, 1986; Kumar, 2007; Sivapalan, 2006), synthesize the current knowledge (Blöschl, 2006; Blöschl et al., 2013), or bring focus on research questions and grand challenges (Blöschl et al., 2019; Montanari et al., 2013; Sivapalan, 2003). Despite the weak use of causal vocabulary in the front end of articles (Figure 2-1), it is between the lines of these initiatives that the importance of causality is acknowledged: "*the exploration of the underlying causal mechanisms, will ultimately become the basic elements of the new theory of hydrology at the basin scale*" (Sivapalan, 2003); "*The reaction of society to change is necessarily related to its causes and therefore an informed quantification is needed*" (Montanari et al., 2013); "[…] *the science of hydrology where the ultimate goal is to understand hydrological causality*"(Blöschl et al., 2019).

In particular, the view of causality has changed through the way the community considers models, that are, causal representations of hydrological systems. Some hydrologists advocate for the reconciling of operational hydrology and the science of hydrology for some sake of consistency between

theory and practice (e.g., Koutsoyiannis, 2014). Complex physically-based models have been subject to significant criticism for decades (Beven, 1989; Grayson et al., 1992; Kirchner, 2006; McDonnell et al., 2007). Even if physically-based, they were increasingly referred to as perceptual, conceptual models, or just hypothesis, agreeing that a compromise must be reached between simplicity and complexity and that new physical laws suitable for the large catchments shall be discovered. On the one hand, the progress in model evaluation and comparison allowed to constraint the jungle by identifying the best models, their parameters and structures, made more comfortable with the development of flexible modeling frameworks (e.g., Clark et al., 2008, 2011; Fenicia et al., 2011; Gupta et al., 2012). On the other hand, simple physically-based models and empirical models have recovered their complementary merits (Hrachowitz and Clark, 2017; Koutsoyiannis, 2016; Savenije, 2009; Todini, 2007). Empirical models are now proposed as a new paradigm to address causal inference in Earth Sciences (Goodwell et al., 2020; Meyfroidt, 2016; Rinderer et al., 2018; Runge et al., 2019b). They were also progressively seen as viable top-down approaches to highlight laws for large catchments, and hydrologists advocates for a merging of bottom-up and top-down approaches (Klemeš, 1983; Sivapalan, 2006; Sivapalan et al., 2003). Today, the top-down approach also focuses on the study of higher-level emergent watershed patterns that arise from a watershed organization (Kumar, 2007; Sivapalan, 2006). As a result, hydrologists have moved away from single-watershed modeling to the comparative analysis of watershed patterns and the development of an appropriate classification framework based on hydrological similarity (e.g., Wagener et al., 2007). Emerging from this task of comparison and classification of populations of watersheds, the Darwinian hydrological approach, by analogy to the evolution theory, asks the question of how some watersheds became what they are (Harman and Troch, 2014).

In general, hydrology has evolved from an engineering or physical science of place to become an Earth science that should be regarded as empirical (Sivapalan, 2018). Some authors go so far as to call hydrology an inexact science (Beven, 2019b; Beven et al., 2020). New branches have emerged aiming to bring hydrology closer to other disciplines, such as ecohydrology (Eagleson, 2002) or sociohydrology (Sivapalan et al., 2012), hence closer to causality addressed from these fields. This evolution is part of a broader change of worldview from a physically and naturally driven hydrology to the inseparability of hydrological, ecological, and human systems at the Anthropocene era, which stresses the importance of interdisciplinarity (Abbott

et al., 2019; Lall, 2014; Montanari et al., 2013; Savenije et al., 2014; Vogel et al., 2015; Wagener et al., 2010). Online platforms have emerged to share data and collaborate from anywhere on Earth (Ceola et al., 2015). Hydrologists works on how best communicate our messages to the general public (Bogaard et al., 2017; Lutz et al., 2018), and how to frame knowledge and skills, despite the everchanging landscape of hydrological sciences, to transfer them to the next generation of hydrologists (see Seibert et al., 2013).

Notwithstanding these great perspectives, hydrologists invest themselves in increasingly challenging duties (Clifford, 2002). Hydrologists keep seeking new hydrological laws, hopefully, simple ones, but now deal with water crisis concerns in various socio-cultural environments (Sivakumar, 2011b) that strengthen the uniqueness of hydrological systems. They shall highlight solutions to the Anthropocene challenges based on future scenarios while dealing with a changing environment, where the past and the law of nature are no longer seen as the best basis to depict the future (Koutsoyiannis, 2013; Milly et al., 2008; Srinivasan et al., 2017; Thompson et al., 2013).

All these concerns reflect most of the epistemological issues that have occurred in hydrology the last 40 years, and they are all, somehow, related to causality. This paper aims to bring the somehow to the how by mirroring the controversial history of causality with epistemological issues in hydrology. In this respect, the approach is similar to Darwinian hydrology (Harman and Troch, 2014). What the causality of hydrological systems is not asked in the first place, but how causality itself has evolved into what it is, and how it connects to the broader history of the concept, coevolved in different fields of science (Figure 2-1) as well as between the various way of doing hydrology. Meaningful concepts related to causality or its siblings as truth, knowledge, or understanding are traced to highlight the multiple facets of causality and potentials for a synthesis (listed in Appendix III). If hydrologists pursuit the cause of a unified understanding, one can indeed argue that the first step is to inquire about the federating concept of causality.

2.2 Sensing Species of Causes

Harman and Troch (2014) pointed out that while Darwin's theory is about evolution, he had at his disposal a taxonomy of species that defines them based on the essential traits that all individual in the same species shares. We will thus start our journey by looking at existing intuitive definitions of the word

"cause" before trying to understand how they came to be. The Merriam-Webster's Learner's Dictionary reports three distinct and straightforward definitions:

(d1) Something or someone that produces an effect, result, or conditions (*cause-effect relationships*, e.g., heavy rainfall causes floods);
(d2) A reason for doing or feeling something (*perceptual causes*, e.g., floods prevent someone from building in risky areas);
(d3) Something (such as an organization, belief, idea, or goal) that a group or people support or fight for (*teleological causes*, e.g., flood risks should be reduced in the future).

These three definitions express three vantage points to address causality. The first one (d1) is instead that of a third-person observer who observes a cause and the effect it produces. The second (d2) is focused on first-person perception, that of a human behavioral agent who feels the world and acts accordingly. The third one (d3) brings a socially shared or constructed context or an idealistic value to the notion of cause.

Most likely, the definition (d1) is the one that natural scientists would rely on, as they are third-person observers analyzing nature. Indeed, definition (d1) expresses causes in the sense of cause-effect relationships, which is the most common meaning in science (Sivapalan and Blöschl, 2017). In general, cause and effect are ordered in time in such a way that cause precedes effect. However, cause and effect can be perceived as instantaneously related (e.g., gravity), just as they can also arise from intrinsic properties or spatial ordering as well. For instance, the cause of water in the liquid phase condition is related to a temporal mechanism such as condensation or melting, or to the molecular and structure of the water and the mechanisms of hydrogen bonds. In that way, some cause refers to the time-invariant internal quality of an object, while some others involve motion, time and occur in a contextual situation. In general, when the cause and the effect are identified and derived from observation only, the causal law is empirical. When such an empirical relationship is understood either from the light of another causal theory, eventually relating motion with intrinsic properties of interacting objects, or if the empirical relationship holds long enough to become a theory itself, it becomes a mechanistic causal law, as the mechanisms are supposed to be unveiled or discovered. The same distinction between intrinsic causes and situational temporal causes holds for perceptual causes (d2): peoples may act

or feel according to their dispositions (e.g., happy/sad, afraid/fearless) or according to their environment (e.g., sunny or rainy). The last definition, with respect to time, depicts a cause to be maintained or pursued, which will eventually be achieved in the future (e.g., a unified theory of hydrology). The temporal ordering is opposite to definition (d1) in that the cause (d3) is after the effects and drives people to the future by supporting a purpose. Still, the term "driver" is also used to qualify a cause of change in a system (e.g., urbanization, deforestation). Not only for people, the term "driver" may also symbolize a goal-directed transition from one state to another in transient natural systems looking for a new end-state or equilibrium. The term teleological comes from the ancient Greek telos meaning both goal or end.

Definition (d2) and (d3) are also noteworthy for scientists. The question is whether these definitions are independently related to causality or whether they create the whole concept through interaction? For example, a scientist who discovers a cause-effect relationship (d1) will most likely support his or her discovery with others (d3) so that other people feel and do things differently (d2). The other question is how these three definitions emerged from the past. As a starting point, we will follow the invitation of Erwin Schrödinger that suggested that the scientific representation of the world consists of a specific attitude of mind, which arose among Greek thinkers and was transmitted to us (Schrödinger, 1954).

2.3 On One Origin of Causality

2.3.1 The Presocratic Revolution: Looking for *Archê*

At the very beginning was chaos, a word that still transcribes the apparent lack of causality in modern science. From this chaos, as described by the poet Hesiod (8th c. B.C), has emerged the first attempt to bring some order: the gods. However, the gods were not a parsimonious model to explain the world, including hydrology, since each river was attached to a god (Brewster, 1997). Two centuries later, the presocratic revolution arises: the gods' fickle characters are no longer considered as sufficient reasons to explain the world (Curd, 2019). Presocratics suggested that the world is causally governed by a few numbers of first principles called *archê*, explaining both the origin of the world and how it has been transformed. The revolution started with Thales of Miletus (6th c. B.C.), the first philosopher, physicist, from *phusis* meaning nature, or even hydrologist (Koutsoyiannis et al., 2007), who recognized water

as the first principle. Far from unanimity, the Greek Antiquity debated the *archê* and their essence for two centuries. Many kinds of models arise from these debates (Curd, 2019). The reductionist one of Democritus (5th c. B.C.) suggests that the world can be explained from tiny indivisible particles, *atomos*, that recall modern physics. The systemic one of Empedocles (5th c. B.C.) considers that the world involves four elements that evolve and get organized based on two driving forces, love and strife, that remind coevolution and survival of the fittest in biology. Some fundamental principles of logic are laid down, such as the principle of non-contradiction embodied in Parmenides' poem (6th c. B.C.), stating that something cannot be true or false at the same time or in the same manner. In summary, the philosophical debate over *archê* revolved around three main concerns: (1) their number: are the first principles one, plural, or infinite?; (2) their substance: are they made of matter, spirit, or both?; and (3) their persistence: are they eternal or changing as suggested by the presocratic Heraclitus (6th c. B.C.) with his "*panta rhei*" (everything flows)?

If reinterpreted through the scope of hydrology, these questions remain surprisingly relevant today, considering (1) the pluralism of hydrological definitions, laws, and models, from their process representations to their representative modeling units (e.g., Beven, 2006b; Blöschl, 2006; Reggiani and Schellekens, 2003), together with (2) the question of the physical realism of these concepts (Beven, 1989; Grayson et al., 1992; Kirchner, 2006), and (3) their persistence in a changing world, where Heraclitus is still quoted today (e.g., Koutsoyiannis, 2013; Milly et al., 2008; Montanari et al., 2013; Wagener et al., 2010).

2.3.2 The Golden Age of Athens: Philosophy, Sciences, and Society

The Athens of Pericles (5th c. B.C.) was the home of many thinkers, three of whom had divergent opinions on what knowledge or causality is: Plato, Protagoras, and Aristotle (Table 2-1). All were philosophers, but given what they could represent today, they are introduced as the philosopher, the manager, and the scientist. Plato's dialogue *Theaetetus* is about the question: what is knowledge? It features the emblematic Socrates discussing with a young man, Theaetetus (see Chappell, 2019).

In a dialectic way, that is, starting by stating different point of views, Theaetetus suggests three definitions of knowledge, as:

(d4) Perception;
(d5) True belief;
(d6) Justified true belief.

The first (d4) instead reflect Protagoras' definition of knowledge embodied in his famous quote: "*Man is the measure of all things*", and is comparable to definition (d2). Socrates rejects this subjective and relativist proposition with many arguments, one of them being that perception can be false about what the thing is, and could, therefore, not be knowledge. The second (d5) is also rejected as a belief is just empirical, e.g., about the fact that the sun will rise tomorrow, but without justification, one cannot assert that the belief is true. This leads to the third definition (d6). Unfortunately, the speakers were unable to agree on what constitutes an appropriate justification. The debate is still up-to-date in modern theories of knowledge (see Ichikawa and Steup, 2018).

Table 2-1: Account of Plato, Aristotle, and Protagoras on knowledge and causality

	Plato **The philosopher**	**Protagoras** **The manager**	**Aristotle** **The scientist**
On Knowledge	*Knowledge is justified true belief*	*Man is the measure of all things*	*We have knowledge of a thing only when we have grasped its cause*
Purpose	Unified understanding for the sake of enlightenment and goodness.	Success in the management of personal and public affairs	Understanding necessary truths and their causes
Method	Philosophy, dialectic, non-contradiction, dialogues, and allegories	Speech, debate, argument, language, and communication	Observation Logic, The Four Causes, Treaties
Grounding	Transcendence World of Ideas	Personal and societal perceptions, values, belief, and norms.	Immanence Real World

In general, Plato gives no clear guidelines to address causality or truth. His thoughts are never developed in first-person treatises but dialogues involving third parties and the crossing of divergent opinions with the dialectical method. So is the *Theaetetus,* where Socrates applies his well-known method

of critical thinking that seeks to refute any claims of his interlocutors in virtue of the principle of non-contradiction, in the spirit of Popper's falsificationism (Popper, 1959). Even more disturbing, Plato portrays Socrates in his *Apology* as the wisest man in the world for his words, "*I neither know nor think that I know*." Still, to Wolfsdorf (2011), the nature of justification in *Theaetetus* can be likened to a broad notion of cause, *aitia* in ancient Greek, that gave birth to etiology, the study of causation. It referred to an answer to the "why" question and was instead developed by Aristotle (section 2.3.3.1).

Answering such a "why" question for Plato is undoubtedly related to his metaphysical and dualistic philosophy. Plato provided a view harmonizing presocratic concerns by recognizing the existence of two worlds (Robinson, 2017): (1) the world of particulars where things are plural, made of matter, but ephemeral and imperfect, and (2) the world of Ideas (or Forms, Universals) made of a few universal principles of an eternal spiritual nature. As Ideas are the purest form of reality, Plato's notion of causality is grounded in the world of Ideas, based on *transcendence* (Delacy, 1939): Ideas govern and infuse themselves into the real world giving rise, essence and motion to particular objects. This view is best embodied in Plato's allegory of the cave (Wheeler, 1997).

Figure 2-2 illustrates the allegory with a caricature of catchment hydrology. From this point of view, hydrologists can be seen as prisoners of the real world, looking at imperfect, seemingly unique watersheds (Beven, 2000), and should work to free themselves, come out of the cave, and find the enlightenment, which would allow them to see and reach the idea of the watershed, a unique and universal understanding of all watersheds in the real world. Although caricatured, the comic illustrates a widespread idealistic definition of science and causality, like the one of Bronowski (1956), "*All science is the search for unity in hidden likenesses [...] order must be discovered and, in a deep sense, it must be created. What we see, as we see it, is mere disorder.* The definition was echoed in hydrology (Klemeš, 1982; Sivapalan and Blöschl, 2017) and is somehow embodied in the quest of universal hydrological laws (Beven, 2006b; Dooge, 1986; Sivapalan, 2006) and watershed blueprints (Beven, 2002b; Freeze and Harlan, 1969; Montanari and Koutsoyiannis, 2012; Savenije and Hrachowitz, 2017). From that point of view, causality itself is a *teleological cause* (d3) driving science, including hydrology, where the ultimate goal of hydrology can be seen as the causality of the water cycle (Blöschl et al., 2019). On the other hand, modelers with

strong beliefs in the causal nature of their model are not that far from Plato's idealism in such a way that there would believe in a causal link, similar to (d1), though, a transcendental one between their model and reality.

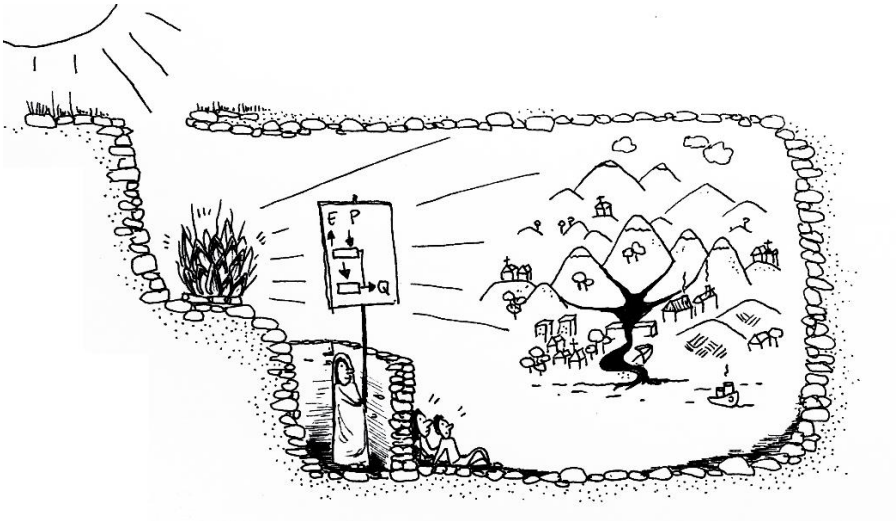

Figure 2-2: The cave allegory and the idealist quest of a unified theory for catchment hydrology (artwork by Morgane Gloux).

Regarding Protagoras, knowledge or causality could be perception (d2, d4). Yet, in his defense, truth is not entirely subjective and relative to individuals, but to the society to which the individual belongs (Taylor and Lee, 2016). Hence, truth and causality are interpreted with respect to a system of shared belief, values in the society (d3) that are embedded in norms, the moral and political laws (*nomos*). Another way to see it is the proper justification of knowledge (d6) is social and societally constructed, or "*created*" in the word Bronowski (1956), but not static as eventually reframed using speeches, debates, with the weapons of language, communication, and argumentation. In doing so, Protagoras is accused of distortion: to make weaker arguments stronger. However, his vision of causality (d2, d4) is not for an end of enlightenment or as mean of understanding, but as a mean of being successful in life, that involves social processes such as perceptions, communication, action, and intervention. From now, the manager's causal vision is a relativist pragmatism rooted in societal grounds.

What about the place of science and the laws of nature in his vision? It seems that Protagoras does not exclude *phusis*-related arguments in decision-making, insofar as it constitutes a genuine argument of authority for the laws, norms, and conventions that are adopted, but the *nomos* prevails on the *phusis* (Boyaci Gülenç, 2016; Taylor and Lee, 2016). Protagoras then values all arguments including scientific ones, as politicians today in the words of Kirchner (2017): "*Science is respected in the policy process because—or, more precisely, to the extent that—it is perceived* [d2, d4] *to convey unbiased, factual information* [d1, d6]". Protagoras' definition of science would instead be based on its role and relevance in responding to societal needs and demands, as sometimes echoed in hydrology (Lall, 2014; Sivapalan and Blöschl, 2017).

On his side, Aristotle's work is somehow pursuing what could be a proper and non-subjective account for definition (d6). Aristotle's legacy is impressive, manifold, and still incredibly up-to-date, and developed within detailed treatises such as modern scientific writing, with many references to the ancient philosophers (Shields, 2008). His corpus divides the different disciplines of knowledge, in this case, drawing a clear line between metaphysics and physics and thus breaking with the idea of transcendent causality (Figure 2-2). In other words, he did not believe that the real world is entirely governed by Ideas and the principle of *transcendence*, but by principles internal to the real world, i.e., by *immanence*. His definition of nature, *phusis*, regroups all things that are ruled by themselves. Thus, physics is the study of all immanent and necessary truths and principles in the real world. Aristotelian nature is opposed to his notion of technique (*technê*) that refers to things designed or driven from the outside, such as human-made objects or arts. To Aristotle, scientific knowledge is a body of organized necessary truths that arises from the study of causes (Angioni, 2016). Necessity is a deterministic notion in the sense that it targets the essence of things, determining what they necessarily are, not on the details that could be referred to as chance, fortune, accident, or randomness. To address what is essential, causal or not, Aristotle had developed the first explicit framework to deal with causality in nature in his *Physics*, and laid down the basis of logic in the *Organon* (meaning tools, instrument). Since these materials are of great importance to understand the evolutionary trail of causality, they are developed specifically in the next subsection.

2.3.3 The Legacy of Aristotle

2.3.3.1 Aristotle's Four Causes

In Aristotle's *Physics*, causality is defined as the capability of answering "why" questions (Falcon, 2019; Shields, 2008). To this end, he suggested a pluralism of four causes that shall be studied to provide a complete explanation and justification for knowledge (d6). His framework is fit for everything, not only for nature, as shown by his example of a bronze statue :

(d7) Material cause: "*that out of which*", e.g., the bronze of a statue;
(d8) Formal cause: "*the form*", "the account of what-is-to-be", e.g., the shape of the statue;
(d9) Efficient cause: "*the primary source of the change or rest*", e.g., the sculptor, the art of bronze-casting the statue;
(d10) Final cause: "*the end, that for the sake of which a thing is done*", e.g., art, representing someone, for a buyer.

In other words, a complete causal explanation to the "why" is nested with four other types of questions: "from what" (d7), "into what" (d8), "how/by" (d9), and "what for" (d10) (for comparison, see Blöschl et al., 2019). The efficient cause (d9) gave birth to the modern notion of cause and effect (d1), the mechanistic one that highlights how the lower-level elements, the materials (d7), turned into a higher level of organization, the form (d8). While studying nature, it is vital to free the causes of human feelings, beliefs, and desires (d2,d4), as they are independent of the immanent nature. A scientist, in line with Aristotle's legacy, will instead approach the efficient cause (d9) of the statue by focusing on the art of bronze-casting, the mechanical recipe, rather than from the sculptor's intentions (Falcon, 2019). Similarly, Aristotle formalized teleological causes (d3) with his final causes (*telos*, d10). However, he did believe that nature had teleological purposes in itself, that is independent of human's (or gods') intentions (d2). For instance, one cannot understand something as simple as an egg by considering its teleological purpose, i.e., its endpoint, or asymptote of giving birth (Grene, 1976).

Aristotle was accused of blurring the canon of causation by relying on different types of causes (Shields, 2008). In addition, Aristotle's causes are lacking guidelines. How do we scale the proper material cause? What characterizes the form of the essence of a watershed? Aristotle would most

likely answer that it is a problem of conceptualization (Klemeš, 1983). In his introduction of *Physics* (*185b8-185b9*, see Aristotle, 1995), he succinctly grasped in one paragraph, the problem of causality concerning closure, continuity, and confounding factors, which sums up the main difficulty of hydrology as a science:

"Now we say that the continuous is one or that the indivisible is one, or things are said to be one, when the account of their essence is one and the same, as liquor and drink. If their One is one in the sense of continuous, it is many; for the continuous is divisible ad infinitum. There is, indeed, a difficulty about the part and whole […], whether the part and the whole are one or more than one, and in what way they can be one or many, and, if they are more than one, in what way they are more than one. [..] Further, if each of the two parts is indivisibly one with the whole, the difficulty arises that they will be indivisibly one with each other also".

To solve the problem of conceptualizing the material or formal cause (d7, d8), Aristotle's final cause (d10) is of great help. Aristotle saw that accident, coincidence, chance, or randomness as an event or fact that lacks final causes (d10), even if they have a material, formal, efficient one (d7, d8, d10). In this way, a default on an eggshell has no place in causal analysis of the egg whose ultimate goal is to focus attention on the material, formal, and efficient aspects that pursue the goal of hatching. This kind of philosophy is referred to as *functionalism*, best depicted by the aphorism echoed in the field of architecture: "*form follows function"*, the rest is unnecessary details. In this way, giving meaning to things, necessarily reduce the dimension of a problem. Still, things may have plural meaning: the final cause of an egg could be eating, in which case the egg is causally framed differently, e.g., by a nutritionist. The growth of knowledge should investigate plural meaning, and Aristotle would have indeed argued that scientist are looking at randomness if they do not link their analysis to a natural purpose giving meaning to nature and science. This view is comparable to the one of Sivapalan (2009) emphasizing the importance of purpose in hydrology, while quoting Seneca: "*If one does not know to which port one is sailing, no wind is favorable"*.

2.3.3.2 The Roots of Logic

Causality is not just about framing or conceptualization. It also involves correct reasoning logic. The purpose of logic is *inference*, that is, identifying

true statements or causal relationships. Aristotle explicitly recognized two fundamental reasoning approaches (Table 2-2): *deduction* (d12) and *induction* (d13). These have only been amended recently by Charles Sanders Peirce (1839-1914 A.C.) with a third form of reasoning, i.e., *abduction* (d14), that we will consider for the sake of completeness. Peirce claimed, however, to have been inspired by Aristotle (Peirce, 1960). In order to have generic definitions, these reasoning approaches are expressed in relation to the concept of *abstraction* (d11) defined as a mental object or ideas within the mind, on which reason can act. For example, hypotheses, definitions, physical laws, or conceptual models are all abstractions because they are the abstract objects of various complexity on which reasoning and logic are based. Typically, induction is the generalization process that creates abstractions from observations. On the other hand, deduction is the logic that manipulates simpler abstractions (hypotheses) to create new and possibly more complex ones (thesis) by rules, such as a mathematical demonstration. Abduction is somehow the opposite of deduction. It seeks to infer the best hypotheses that support some observed and induced abstractions, or facts, such as the work of a detective. Abductive logic was explicitly introduced very recently by Baker (2017) in hydrology in the light of recent debates on hypothesis testing (also discussed in Harman and Troch, 2014). Imagination (d15), *phantasia* in Aristotle's *de Anima* (Shields, 2016), is the last component we choose to include, without discussing it, but for the sake of closure and to not reduce reasoning to the strict formality of logic. In this way, it is permissible to consider that not all abstract ideas come exclusively from induction (d13), but from elsewhere or spontaneously, as a bright or crazy idea, which may be necessary for advancing science (Burt and McDonnell, 2015; Klemeš, 1983; Sivapalan and Blöschl, 2017).

Table 2-2: Abstraction, the three fundamental reasoning approaches, and imagination

Label	Term	Definition	Example
(d11)	Abstraction	An abstract mental object	All of the examples below.
(d12)	Deduction	Inference in which a new abstraction (thesis) about particulars follows necessarily from prior general or universal abstractions (hypotheses)	Since rain falls, the ground will be wet.
(d13)	Induction	Inference of a generalized abstraction from the sensory observation of particular instances in the real world.	Rain is always followed by the ground being wet; the fact that rain precipitates; what rain itself is in general.
(d14)	Abduction	Inference to the most likely general and universal premises associated with a conclusion generalized from particular instances.	The ground is wet, it must have rained
(d15)	Imagination	The mental ability to create abstractions not exclusively based on sensed particular instances	An equation; a virtual abstract reservoir; an elephant watering the ground.

2.3.3.3 Actuality and Potentiality

Understanding the cause of a static object as a statue is an easy task compared to the analysis of an everchanging object such as a watershed. Aristotle laid down two other principles in his *Metaphysics* (Cohen, 2016; Marmodoro, 2018), also discussed in his *Physics* (Book III), that are convenient to address and summarize the frustrating and widely debated Heraclitus' problem (Koutsoyiannis, 2013; Koutsoyiannis and Montanari, 2015; Lins and Cohn, 2011; Matalas, 2012; Milly et al., 2008, 2018; Montanari and Koutsoyiannis, 2014) of finding some stationary causes when things change and involve motion:

(d16) Actuality (*entelecheia* or *energeia*): what is capable of being seen, the account of what is seen, e.g., the statue of bronze, the artisan crafting the statue, data;

(d17) Potentiality (*dunamis*) or dispositionality: what is capable of being built, the capacity to be in a different state, the account of what-could-be, e.g., a sword of bronze, forecasts.

Therefore, actuality is preferentially related to the induction (d13) and potentiality to the deduction (d12). The efficient cause or processes (d9), e.g., the art of bronze casting, is supposed to give an account of what-could-be (d17) by deduction, similar to a hydrological model running probable scenarios (Montanari and Koutsoyiannis, 2014). To Aristotle, a transition from potentiality to actuality is metaphysical causation, similar to Figure 2-2. In other words, the law transcends the real world. Hydrological models, physical law, a fitted statistical model, are somehow all expressions (d9) of potentialities that are a static representation of motion (d8) generalized by induction (d13). They allow reasoning by deduction (d12) to extrapolate the field of what-is-to-be (d8) to the what-could-be (d17). However, Aristotle had understood the *problem of induction* (section 2.4.3.1), and why such models induced from observations may be wrong (*Phys. 201b24-202a2*): "*motion is thought to be a sort of actuality, but incomplete, the reason for this view being that the potential, whose actuality it is is incomplete*". Failure of model predictions, or engineered infrastructure, can, therefore, be seen as a wrong generalization of potentialities due to incompleteness of data. Some see it as a change of potentiality, using the term nonstationarity (Milly et al., 2008). Some others call it change, just a perceived one revealed by new actuality, that allows us to correct our underframed or fitted account of potentiality, that has in the end ever been stationary or time-invariant (Koutsoyiannis, 2006a, 2013; Lins and Cohn, 2011; Montanari and Koutsoyiannis, 2014), and that forces hydrologists to incorporate the new agent of change (Ehret et al., 2014; Montanari et al., 2013; Troch et al., 2015).

2.4 Hydrology Through the Evolutionary Path of Causality

2.4.1 Middle Ages: Is Hydrology Realistic?

For the sake of brevity, causality related thoughts belonging to the long period of the Middle Ages won't be developed, in spite of genuine opinions on causality (see White, 2018). Generally speaking, the Middle Ages were dominated by Platonic duality between Particulars and Forms, matter and spirits, or body and souls (Spade, 2018). It is only in the 13th century that Aristotle was reintroduced in Europe, notably, by St Thomas Aquinas (1225-1274 A.C.). Both causality and Platonic dualism were reinterpreted through the prism of the Judeo-Christian faith, mostly by scholastic philosophy. Concerns of causality were instead focused on free will (d2) and the

explanation of motion, i.e., on the efficient causes (d9) that became simply termed causes. Indeed, at that time, the material cause (d7) and the formal one (d8) lost their status of causes to became instead explanations or "because" (White, 2018). The final or teleological cause (d10) took on a cosmic meaning and is attached to the designs and plans of the Judeo-Christian God (Grene, 1976).

Still, one cannot speak about causation without considering the controversial theses of Ockham (1285-1347 A.D.) that reject the existence of universal concepts, e.g., Platonic Ideas (Spade and Panaccio, 2019). By that, he meant that universal concepts are mental abstractions (d11) of reality and not reality itself. In this respect, the controversy launched by Ockham, i.e., the *problem of universal*s, remains relevant in hydrology. Indeed, Figure 2-2 can be reinterpreted as there may be two caricatural and opposite views within the hydrological community: one that believes in a universal representation of catchment hydrology (Clark et al., 2015; Dooge, 1986; Sivapalan, 2006), and those who look at catchments as unique entities (Beven, 2000; Blöschl, 2006). Ockham's debate also echoes to Box's famous quote (Box, 1976): "*all models are wrong, but some are useful*". Ockham goes further than such an aphorism and tells us how universal ideas, even though false, can be useful: they are economical and allow us to explain a variety of concrete real-world objects under the same term, under a single abstract representation (d11). This metaphysical belief (d2, d3, d5) is known as:

(d18) The principle of parsimony, or Ockham's razor: one should not multiply reasons without necessity (Baker, 2016).

The principle is regularly considered in the evaluation and choice of models, sometimes explicitly relating the performance of a stochastic model to its degrees of freedom (Akaike, 1974; Schwarz, 1978). Likewise, in hydrology, the search for simplicity and parsimony remains a concern (Dooge, 1997; Koutsoyiannis, 2016; McDonnell et al., 2007; Weijs and Ruddell, 2020). The question is: can we elaborate a parsimonious process-based (d9) model suitable for all catchments? In the current state of hydrology, generically, universal models are complex (i.e., sophisticated, section 1.1.2.1) but not parsimonious, while parsimonious ones are not universal. Still, the new emphasis on catchment classification (e.g., Wagener et al., 2007) expects to develop a taxonomy based on hydrological similarity, such that each watershed in a taxon could eventually be modeled parsimoniously. However,

if a universal model or framework is recognized as a good teleological cause (d3) driving the progress of the science of hydrology, many hydrologist belief that there will still be an actual need for a plurality of models and types of models (either complex or simple), if not because of catchment uniqueness, at least to address particular practical applications (Weiler and Beven, 2015), i.e., to frame the model around what for it should be useful (d2, d10). The problem of universal is still an open debate.

2.4.2 Rationalism: Could Hydrology Be a Machine?

Rationalism emerged with Descartes (1596-1650 A.C.), among others, in a period where the Church had a dogmatic authority over some domains of knowledge and where scientists such as Galileo were prosecuted for their controversial causes (d3). In this context, Descartes claimed that knowledge should go back to basics: pure reason and the use of deduction (d12), relying on undoubtful truth as premises that have passed the test of Descartes' methodological doubt (Descartes and Renault, 2016). His scientific method suggests addressing complex problems by splitting them into simpler problems and then to reconstruct complexity from simplicity. In the same vein, Descartes is also attributed to the view of the animal-machine, the human body (d8) being like a machine and could be mechanistically described (d9) from its parts (d7). This mechanistic worldview is nowadays referred to as bottom-up or upward causation (Klemeš, 1983), since the higher level, the whole, is explained from the lower-level components. Another rationalist, Leibniz (1646-1716 A.C.) formalized a vital principle regarding cause-effect relationships (d1, d9), that can be explained by posing the distinction between two other types of causes:

(d19) Principle of sufficient reason: everything has a cause (Melamed and Lin, 2018);
(d20) Necessary cause: a cause/reason/condition that is always involved in the realization/observation of an event/state;
(d21) Sufficient cause: a cause/reason/condition or a set of it that always implies the realization/observation of an event/state.

Hence, Leibniz's principle (d19) says that for all events in the world, there exist causes that explain them, suggesting that these causes will always imply the event (d21). The principle reflects a deterministic worldview that is emblematic of this rationalist period. At the same time, the principle is a strong

impulse for causality, stating that everything is worth a partial (d20) or complete (d21) causal explanation. Besides, Leibniz contributed to the development of the mathematics of differential equations. So did Newton (1642-1727 A.C.) in the same period. Since then, dynamical systems' states are conveniently explained by a set of differential equations, a mathematical expression of Aristotle's efficient cause (d9) that explain motion (d8, d16, d17) from initial conditions and physical properties (d7). As Newton developed the Newtonian mechanics, the mechanistic, physically-based, or dynamic hydrology (Eagleson, 1970) is also often referred to as Newtonian hydrology (see Sivapalan, 2018).

The merits of rationalism and mechanistic causation (d1, d9) cannot be denied. Most of the progress in modern physics and the advance in technological development comes from this philosophy. Although the approach is sometimes described as reductionist, Klemeš (1982), quoting Ziman (1978), reminds that "*whatever one's philosophical attitude towards reductionism, there is an inescapable scientific necessity of trying to 'understand' and 'explain'* [d8, d9, d10] *the behavior of any system* [d9, d16] *in terms of a relatively few comprehensible elements* [d7] *without recourse to an elaborate extracerebral computation* [d18]*".* Hence, all causal models are reductionist in that sense of reduction to the parts (d7, d9) or in whatsoever sense, since it will always reduce causal reality to an abstraction of it (d11). Also, deduction (d12) is the only logical process that allows to reason outside of the realm of experience, which is a necessary condition (d20) for technological development and to make a forecast in the future under changing conditions (section 2.3.3.3). This capability of extrapolation (d12, d17) by consideration of the mechanism (d9) is the primary motivation and asset of the physically-based model since the beginning of their development (Freeze and Harlan, 1969).

The main fallacies of deduction (d12), as acknowledged by Descartes, are easily summarized: deduction is the most robust type of inference, but it derails when the reasoning is built from imperfect, uncertain, or wrong pieces of knowledge as hypotheses. In that regard, mechanistic determinism was rapidly challenged by the apparent randomness of nature, which was seen as a human failure in seeing sufficient causes (d21). Laplace (1749-1827 A.C), with his well-known demon, suggested that the world (d8) can still be mechanistically apprehended (d9), but by a supreme intelligence that would be able to perfectly grasp the state of the universe and the forces at work (d7)

(Laplace, 1814). This fantasy (d15) supports the dream of science, including hydrology, that will free itself from the empirical need of data (see Silberstein, 2006). Often portrayed as an emblem of causal determinism (d9, d19), Laplace was nevertheless a pioneer formalizing the logic of inference by induction (d13) found in the theory of probability that has emerged from the work of Bayes (1702-1761 A.C.). Since the human mind is far from being a supreme intelligence, Laplace motivated his essay on probability by the need to grasp the apparent randomness of the world due to our lack of knowledge. This lack of knowledge is often referred to as epistemic uncertainty (e.g., Beven, 2016). From the mechanistic point of view, which became the paradigm of physically-based hydrology, randomness is sometimes seen as the "evil" and determinism as the "good" (Koutsoyiannis, 2010), and hydrologist should work to reduce randomness and uncertainties. This view is very different from Aristotle's one that suggests that randomness is lack-of-purpose noise (d10, section 2.3.3.1). Regarding uncertainties, Nearing et al. (2016) reminded the extent of many discussions on the role of uncertainties (d3, d10; e.g., Pappenberger and Beven, 2006; Sivapalan, 2009), their nature (d7, d8; e.g., Koutsoyiannis, 2010; Montanari, 2007), and their appropriate handling (d9; Beven et al., 2012, 2008; Clark et al., 2012; Mantovan and Todini, 2006; Stedinger et al., 2008; Vrugt et al., 2009). Thus, although fuzzy by nature, randomness can be studied as well in the light of Aristotelean causality by asking its source (d7), its form (d8), its propagation (d9), as well as its meaning and purpose (d10).

In phase with Laplace's concern, physically-based hydrological models were deemed imperfect because of their lack of resolution and thus failed to account for the necessary spatial and temporal heterogeneities (d7, d8, d20) that control hydrological processes (d9). In the opposite way of parsimony (d18), this suggests that model faithfulness is correlated with model complexity (Grayson et al., 1992). With such a state of mind, it is expected that the hydrological behavior of a watershed (d8) will be solved automatically and sufficiently (d21) by relying only on the mechanism (d9, d17) with detailed enough data (d7) (see Sivapalan, 2018).

This bottom-up roadmap has been thoroughly criticized. First, obviously, no physically-based hydrological models have ever been able to rely on the sole logic of deduction (d12), such as the law of physics. They are enslaved to calibration (d13). More reductionism in terms of details had the opposite effect, as highlighted by the modeling paradox of model complexity (Beven,

2012a; see also section 1.1.2.2). If the finer resolution and process representation in hydrological models is motivated on a deductive argument (d12), the increasing complexity resulting from discretization makes hydrological models more strongly dependent on the inductive reasoning of calibration (d13), because of the higher number of parameters, and eventually more uncertain if not constraint enough due to equifinality issues (Beven, 2006a). Equifinality refers to the plurality of many different competing model structures and parameter sets within a model (d9) that might give equally acceptable results (d8, d17). Accordingly, these acceptable results may not include a single realistic configuration given that the goodness of fit is granted by the degrees of freedom of the model (Kirchner, 2006). From that perspective, mechanistic physically-based models may be right, especially if the formal cause of the watershed hydrology (d8) is exclusively evaluated on the paradigm of hydrograph fitting (see Sivapalan, 2018; Woods, 2002), but for the wrong reason, as being based on wrong hypotheses that suggest that processes studied in laboratory or hillslopes remain valid for real and large catchment (Beven, 1989). To Savenije (2001) as for many others (McDonnell et al., 2007; Sivakumar, 2008; Sivapalan, 2006), the curse of equifinality is instead a blessing giving some empirical (section 2.4.3) justification to the principle of parsimony (d18), beyond its metaphysical status, that motivate a functional modeling approach (section 2.3.3.1) meaning that the form of the model (d8) follows its function, that is to be "fit-to-purpose" (d2, d3, d10; Beven, 2012a).

Another argument in phase with the same conclusion is that heterogeneities never disappear (Sivapalan, 2018), or the *continuous is divisible ad infinitum* as said by Aristotle (section 2.3.3.1). In pursuing this way, hydrologists will inevitably end up with hydrological models operating at the scale of subatomic particles. This non-sense was discussed by Klemeš (1982) that advanced that the purposes of fragmenting science "*is to split the (possibly) infinite causal chain into segments, each containing only a few chain links; the scope of one discipline is thus intentionally limited to seeking causal relationships* [d1, d9, d17] *among phenomena* [d8, d13, d16] *within only a relatively small range* [d18], *whose lower boundary* [d7] *represents the discipline's first principles or scientific basis coinciding with the discipline's objective* [d3, d10]*"*. He later referred to this task succinctly under the term of conceptualization (Klemeš, 1983), which appears to be a critic reminder that hydrologists shall think about the material, the formal causes, and their scales (d7, d8), as they will not be solved from themselves by encoding hydrological processes (d9)

in a computer program. This has motivated (d2) many researchers to focus on those lower boundary as representative elements for watersheds (Beven, 2006b; Reggiani and Schellekens, 2003). Conceptualization remains a tedious task (section 2.3.3.1) due to the range of scales, both spatial and temporal, involved in the hydrological cycle (see also Blöschl and Sivapalan, 1995).

2.4.3 Empiricism: Hydrology Beyond Association and Refutation?

2.4.3.1 Model Evaluation: the Problem of Induction

Despite his connection to rationalism, Newton said about his theories, "*hypotheses non fringo*", i.e., *I do not make hypotheses*, stressing the empirical foundations of all sciences. In fact, distinguishing empirical from mechanistic law is very challenging. To be fully causal (Klemeš, 1982), a model should operate without recourse to calibration. Then, even when based on physics (section 2.4.2), empiricism is intrinsic to hydrology, and all models are empirical. From this point of view, the only criterion that could distinguish an empirical model from a physical-based model is through its ability to remain correct under hydrological conditions different from those in which the model was developed (section 2.3.3.3; 2.4.2; d9, d16, d17). This is how models are in practice cross-validated with the split sample test routine (Klemeš, 1986b). However, many empirical models, including physically-based models working for the wrong reasons, can succeed and be operationally used as long as the system under study is not subject to previously unobserved change. Thus, a change would be required to decide on the causal nature of the model, which will remain ultimately problematic when models aim precisely at anticipating or predicting the nature of this change in order to act and make decisions. Hence, it is not possible to know ultimately whether a model is causal a priori, which relays hydrological predictions to the rank of beliefs (d5), or prophecy (Beven, 1993). This is precisely the observation made by one of the fathers of empiricism and philosopher of causality, David Hume (1711-1776 A.C.), through his shock sentence targeting the astral bodies that Newton had just described: "We cannot know if the sun will rise tomorrow".

What can we know? This is the question of interest to Hume (1738, 1748). As an empiricist, he suggested that the origin of all ideas, as well as the association of ideas and causality, is the result of the synthesis of our sensitive and lived experiences (d16), i.e., *induction* (d13). Hume examined a simple

statement, that is, "a cause produces an effect" (d1), and tried to investigate the sufficient conditions (d21) to assert such a claim. Hume offers four necessary (d20) principles illustrated in Table 2-3. First, the cause must occur before effect (d22). This criterion is also known as the *post hoc ergo propter hoc* adage meaning "after this, thus because of this". The principle of priority establishes the importance of the temporal dimension and time asymmetry regarding causality (e.g., Koutsoyiannis, 2019). Secondly, the cause and the effect coexist on a space-time continuum offering the opportunity for interactions (d23). Thirdly, *constant conjunction* (d24) shall be observed between the occurrence of cause and effect. The principle should be refined by the addendum "under similar conditions" to account for nonlinear effects. Constant conjunction to the same conclusions is also what is expected from independent scientific experiments and thus relate to the concept of the repeatability and reproducibility of scientific outputs. The fourth principle, the *necessary connection* (d25), is surprising because it states that the first three, although necessary (d20), are not sufficient (d21), and requires a mysterious additional necessary connection reminding the difficulty of finding a proper justification (d6) in the *Theaetetus* problem (section 2.3.2). The fourth principle is at the origin of Humean doubt and skepticism, which is a moderate skepticism that does not prohibit speculation with enthusiasm on causality, or getting close to it, with, nevertheless, a guardrail reminding us that causality is a belief (d5) and that what is deemed causal today may be deceiving and wrong tomorrow.

Hume's *problem of induction* consists of a wrong empirical synthesis (d13) on the available observables (d16). The problem seems unsolvable since experiencing is always finite in time. A corollary comes out from the logic of Popper (1959): any verified causal claim (d6) is either true or false. Then, the statement was a strong motivation (d2) to change the paradigm of model evaluation from verificationism to falsificationism as one can only assert that something is false. This depicts a definition of truth, and by extension of causality, in the negative. What is scientifically true or causal (d5) is what could be tested but has not yet been proven to be false. One way to address the problem is to seek from new observables (d16) that refute a theory or a model by integrating the technologies of our time (see Peters-Lidard et al., 2017), or previous knowledge to test hypotheses in virtual experiments (Schalge et al., 2016).

Table 2-3: Hume's causal criteria applied to the "rainfall causes river flood peaks" association

Label	Principle	Definition	Example of the rainfall-flood peaks association
(d22)	Priority	The cause occurs before the effect.	Rainfall occurs before flood peaks
(d23)	Contiguity	The cause and effect are contiguous in time and space.	The observed rainfall and flood peaks are closely related in space and time.
(d24)	Constant conjunction	The occurrence of the cause systematically implies the occurrence of the effect [under the same conditions].	Similar rainfall always implies the observation of similar flow peaks under similar conditions.
(d25)	Necessary connection	The additional principle that is necessary to avoid being deceived by the first three.	We cannot know ultimately, but we can speculate infinitely deeper and deeper on overland flow, subsurface flow, hydrological connectivity, etc.

However, Popper's assertion on falsification as the driver (d3, d10) of progress in science, and how (d9) falsification should be pursued, is the subject of recent debates in hydrology (Beven, 2018; Blöschl, 2017). Most agreed that more emphasis on hypothesis testing should be beneficial and adopted using model evaluation framework, field exploration, and data collection. However, it is recognized as difficult in practice as a matter of testability (see also Dooge, 1986). Models are more complex than Hume's focus on one cause producing one effect (d1). They are high dimensional hypotheses themselves (d8) built upon imperfect or incomplete data (d7, d16) that shall be regarded as hypotheses (d11) as well. In conclusion, refutation is necessary (d20) but not sufficient (d21) and shall be pursued in parallel with model development.

It is also illusory to completely discredit the verificationist and pragmatic argument if new technologies make it possible to make better predictions, as could be the case with the new generation of empirical models resulting from the learning machine (Shen et al., 2018). When Mandelbrot's fractional noise model was criticized for being nonrealistic physically, the latter responded that his model is at least empirically more justified, and that all justifications of models of nature are empirical (Mandelbrot, 1970). By the words of Richard Dawkins, when asked the question of the foundations of the justification of scientific truths (d6), the answer is "it works ...". The pragmatic argument is undeniable and eventually more useful than a physically-based model for practical purposes (see also d26). Nevertheless, it does not deprive

scientists of asking why it works and whether it is for the right reason by opening the black box and turning it into a grey box that gives an intelligible (d18) account of the mechanism (d7, d8, d9)(Kirchner, 2006).

2.4.3.2 The Quest of a Necessary Connection

Beyond refutation, another and less pessimistic path is to not give up on the human ability to infer more robust associations, to mine more in-depth the idea of the necessary connection (d25), and to hunt the association-causation fallacies. Accordingly, another empiricist, Berkeley, suggested a precautionary principle, which is often translated in modernity by "*association (or correlation)* [d5,d8, d16, d24] *is not causation* [d6, d9, d17, d25]". In his words (Berkeley, 1710), the transition from association to causality must be made with rules and wise contrivance. The same caution applies today in hydrology (Christofides and Koutsoyiannis, 2011). If Hume is probably responsible for the reluctance of referring to causality in science according to Russell (1912), it is at the end of the 19th century that causality started its come-back, and precisely by looking for associations. With the rise of statistics in the late 19th century, the terms 'regression' and 'correlation' (d24) were coined by Francis Galton in 1888, first termed statistical scales, and later formalized by Pearson through its well-known coefficient as an objective and statistical measure of association (see Stigler, 1989). Accordingly, Yule developed the foundations of time-series analysis: partial correlations (Yule, 1907) and serial correlations that study time-dependencies (Yule, 1921). Cross-correlations or correlograms allows investigating the principle of priority (d22) and identifying causal delays. Besides, building upon correlation and illustrated on agrometeorological variables, Wright (1921) laid the foundations of the modern causal graph theory, now based on Bayesian network, structural equation modeling, among others (see Pearl, 2000; Spirtes et al., 1993).

In the '60s, causality is explicitly back. Reichenbach (1956) stated why an association may causally spurious while verifying Hume's three criteria (Table 2-3, d22, d23, d24):

(d26) Principle of the common cause: constant conjunction (d24) is either the product of causal interrelation or the result of shared driving variables.

Accordingly, two renowned scientists, the mathematician Wiener (1956), followed by the econometrician Granger (1969), proposed an empirical definition of causality based on the principles of priority (d22), common cause (d26), and predictability (d8): a variable is causal to a predicted variable if not considering its past (d22) significantly increases the uncertainty of the predictions while considering all other potential explaining variables (d26). The *Granger causality* is tested with a multivariate linear framework around vector autoregressive models or cross-spectral methods (Granger, 1969). More recently, they have been some development in addressing nonlinear causal interactions with new bi-variate methods based on the information theory (Shannon, 1948) known as transfer entropy (Schreiber, 2000), or on the basis of chaos theory (Sugihara et al., 2012; Ye et al., 2015). The idea is now extended to multivariate nonlinear frameworks (d26) of time-series analysis based on conditional mutual information (Hlaváčková-Schindler et al., 2007; Runge et al., 2019a). A broader view of the evolution, nature, opportunities, and challenges of what is now called a causal revolution or new paradigm can be obtained from some reviews in the field of Earth sciences or hydrology (Goodwell et al., 2020; Meyfroidt, 2016; Rinderer et al., 2018; Runge et al., 2019b; see also Chapter 6).

In general, all these models, that are now labeled causal, are generated from a few time-series. They are non-reductionist (to the small parts, d7) and could be regrouped within a manipulabilist, interventionist, or counterfactual philosophy or theory of causation (Menzies and Beebee, 2001; Woodward, 2016). They are empirical as they tell us what change in one observed quantity corresponds to a change in another (Klemeš, 1982). However, what is implicitly expected with a causally inferred model is based on a pragmatic and intuitive definition of causes:

(d27) Manipulable cause: handling devices to manipulate the effects.

Accordingly, an interventionist would critic a high-resolution model because they are framed upon diluted causes that have lost their status since intervening on a model unit will not much affect the outcome. In hydrology, such a non-reductionist model was often labeled top-down or data-driven if built with the inductive logic (d13) of regression. Even though these methods have had considerable credence in addressing causality in other disciplines, notably in econometrics and neurosciences, time-series analysis techniques or stochastic hydrological models (e.g., Yevjevich, 1987) were not specially

considered as "causal" within the science of hydrology. They were instead labeled empirical, assuming that they miss the causes in the sense of mechanism (Klemeš, 1982). This view was not entirely supported (e.g., Koutsoyiannis, 2010), and empirical causal modeling was developed within hydrology by using similar methods based on structural equations and labeled as data-based mechanistic models (see Young, 2006, 2013). Aside from data-based mechanistic modeling, study cases relying explicitly on causal inference methods are somewhat rare and recent. Some papers investigate potential feedbacks of soil moisture on precipitation and the importance of confounding (d26) factors (interannual variability, seasonality, rainfall-induced synchrony) using Granger causality (Salvucci et al., 2002; Tuttle and Salvucci, 2017). Molini et al. (2010) studied cross-scales rainfall interactions relying on wavelet analysis to address the principle of priority in both the temporal and the frequency domains. In the last decade, nonlinear causal inference methods have been applied to study ecohydrologic feedback processes (Ruddell and Kumar, 2009) and to the study of hydrological connectivity (Rinderer et al., 2018; Sendrowski and Passalacqua, 2017).

In particular, hydrological connectivity is intriguing in the way it echoes Hume's necessary connection (d25). Hydrological connectivity studies the preferential and actual flow paths of water (see Bracken et al., 2013). Following Rinderer et al.'s terminology (2018), structural connectivity refers to the potential ones (d17, d23), independently of water, and solely based on static geomorphology. Functional connectivity, in environmental sciences, refers to the spatial adjacency or contiguity characteristics interacting with temporally varying factors to lead to the connected flow of material (d8, d16, d23), eventually revealed by statistical associations. Rinderer et al. (2018) coined the termed effective connectivity to refer to the connections revealed by causal inference methods, thus, satisfying priority (d22), time asymmetry, and constant conjunction (d24), and depending on the methods, a constraint on confounding factors (d26). Their results also showed that the retrieved causal relationships are method-sensitive and that spurious and physically unrealistic causal relationships may appear unless physically constrained with structural and functional connectivity measures (d16, d17, d23). They further suggested to systematically use multiple methods to infer causality from an abductive logic (d14), and to retrieve parsimonious (d18) connectivity maps by applying a threshold on the significance level of causal associations.

In general, empirical methods are criticized as they merely reveal causal interactions without any explanation of the mechanism. In that sense, an association is just the expected account of what is to be (d8) that could either be an out-of-purpose anomaly (d10) if it does not characterize the whole system (d8). To Meyfroidt (2016), such methods are complementary to qualitative and narrative studies that pay more attention to the causal mechanism. To Runge et al. (2019), causal inference and physical models should be studied complementarily. Moreover, to Rinderer et al. (2018), information about causal and physical mechanisms should be used to constrain causal inference methods. Somehow, in many ways, Hume's necessary connection as a mysterious empirical criterion turns out to be an account of the mechanism (d9). The same conclusion can be drawn from the past evolution of hydrology. Closely related to the cross-correlation method, the unit hydrograph method (Sherman, 1932a) allows addressing priority (d22) and characterize transit or transfer times (d8) between rainfall and river runoff. Sherman (1932b) succeeded in relating (d8) the unit hydrograph to the catchment characteristics (d7, d23, d24), and his method became part of the linear theory of hydrological systems (d9) (Dooge, 1973, 1968).

However, the fact that different empirical causal inference methods outcome different causal structures (e.g., Rinderer et al., 2018) is a recall that each method is carrying its hypotheses: general ones such as general assumptions of linearity or not, structural ones regarding parametric methods, and methodological ones regarding nonparametric methods. From this point of view, these analytical methods are sensitive and, in a way, are a vector of information on the mechanisms. They are not purely inductive and are sometimes very assimilable to a parsimonious mechanistic model. A sharp distinction between empirical and mechanistic remains highly debatable but arguably not productive. In any case, hiding behind the empirical myth of "*hypotheses non fringo*" would deprive us of an opportunity of understanding because they are hypotheses to be explored. Perhaps, Newton hid behind this myth to avoid (d2) being prosecuted like Galileo by dogmatic authorities? Moreover, even data should be seen as a hypothesis (d11) (Beven, 2018) that could be intentionally rearranged and conceptualized in many ways (d7) clustered by expert knowledge (d7, d8, d9), association metrics (d8), or creativity (d15). A fortiori, the human observation was found to be full of tacit hypotheses that shall be explored: "*we know more than we can tell"* (Polanyi and Sen, 1966). All of these concerns relate to phenomenology.

2.4.4 Phenomenology: Hydrology Beyond Perceptions?

Hume's account of causality as belief has influenced Kant (1724-1804 A.C.) and his revolution built on this interpretation of causality: a mental representation of the real world (Berry, 1982; De Pierris and Friedman, 2018). This change of worldview involves a notion of causality that is no longer based on the principle of nature itself (immanence, section 2.3.3.1), but on how nature appears to us through our interpretative and directed consciousness, i.e., on the concept of *phenomenology* (Smith, 2018) developed latter by Husserl (1859-1938 A.C.). From the phenomenological point of view, the mind is not third-person reasoning about the reality of concrete objects in the real world (d1), but a first-person interpreting *phenomenon* as it appears within the mind (d2). Heidegger (1889-1976 A.C.) further stated that the phenomena are a function of the historical antecedent of the individual mind, such that causality is no more absolute but relative to individual minds, groups, or societies. After two millennia, the vision of the sophist Protagoras (section 2.3.2) is back (Caston, 2019; see also Klemeš, 1983): "*man is the measure of all things"*. Consequently, a phenomenological consideration of causality questions the human observer and how he/she takes the measure of things instead of the observed object in itself. Phenomenology, as a new ground for causality, is focused on perceptions and exported itself in various interrelated disciplines: sociology, cognitive sciences, psychology, behavioral sciences or linguistics (e.g., Berger et al., 1967; Brown and Fish, 1983; Heider, 1958; Kelley, 1973).

2.4.4.1 Sensing Causal Perceptions in Society

A phenomenological approach also has advantages in water resource management as well as in the science of hydrology. The subject remains difficult to deal with exhaustively because of the vast field of perceptions related to water. On a blue planet, water is many things to many people: a source of wealth as an essential need for humans and ecosystems, a factor related to the production of food, energy, and goods, a recreational luxury, or a part of our spiritual life. Water is also a source of threats and hazards through excess, contamination, or scarcity. Water is an essential subject to study, understand, manage, legislate, requiring communication across a wide range of media, within and between many different societies, cultures, groups, or individuals. Vogel et al. (2015) point out that water resource management has historically been built on an interdisciplinary basis combining hydrological,

economic, and engineering considerations, as well as legal, political, religious, and ethical considerations (see also Sivakumar, 2011a). Water governance policies have not integrated all water-related concerns at once. Savenije et al. (2014) show how perceptions in water control and governance have changed over time. Initially focused on its hydraulic mission, water governance has developed on a Cartesian reductionist paradigm (Falkenmark, 2001), which has resulted in a bureaucratic sectorization offering mainly technocratic solutions focused on water security. Today, the paradigm is somewhat integrated, adaptive, and systemic (see section 2.4.5). It recognizes the interrelationships between water, ecosystems, and human activities, the multi-purpose role of water, and the plural reality of stakeholders' perceptions through participatory processes in water governance. Noticeably, Falkenmark (2001) quoted the words of the Secretary-General of the European Environment Agency: "*Facts are facts, but perceptions are reality*".

Of these perceptions, which often become perceived causes and turn into action (d2, d27), Table 2-4 summarizes various concepts related to perceived causation, sometimes termed differently. The first six (d28 to d33) are part of a theory of causal attribution (Heider, 1958; Kelley, 1973; Weiner et al., 1987) that studies the naive (and usually biased) way in which people attribute causes to people's behaviors or external events, and how it affects their chances of success or failure. In the same vein, psychology and cognitive sciences have studied the nature of individual behaviors dealing with uncertain risks (Slovic, 1987; Tversky and Kahneman, 1974). These theories sensing people's perceptions have been widely applied to the study of people's behavior concerning natural hazards, such as floods for example (Boholm, 1998, 2009; Buchecker et al., 2013; Burn, 1999; Fuchs et al., 2017; Kellens et al., 2013; Terpstra, 2011). The term causal frame (d34) is widely used and denotes how people, media, politics perceive reality (e.g., Chong and Druckman, 2007; Lakoff, 2010; Tversky and Kahneman, 1981). Frames are seen as heuristics for interpreting reality, understanding, judging, making decisions, and acting in everyday life. In general, people, politics or media exhibit a tendency to have simple biased frames and to attribute a fact or an event to one single cause (Kelley, 1973; Lakoff, 2010; Mackie, 1965), that is often necessary but insufficient (d20, d21), e.g., "rain causes floods". Such a simple causal structure is sometimes referred to as linear or direct causation (or thinking), by opposition to systemic causation (Lakoff, 2010).

Table 2-4: A review of some psychological, cognitive, or sociological concepts related to causality

Label	Term	Definition	Example (for a flood event)
(d28)	Dispositional cause	A cause that is attributed internally to one's personal trait (d2, >< d29)	The manager is incompetent.
(d29)	Situational cause	A cause that is attributed externally to the environment (d2, >< d28)	The manager was overwhelmed; It was an extreme rainfall.
(d30)	Stable cause	A cause that is perceived as temporally persistent (d2, >< d31)	The manager incompetence (if deemed persistent); heavy rainfall and floods always happen.
(d31)	Unstable cause	A cause that is perceived as being temporary or rare (d2, ><d30)	The manager was inexperienced; It has rained a lot the past few weeks.
(d32)	Controllable cause	A cause that is perceived as being manipulable (d2, d27, >< d33)	The manager can learn, rain or urbanism (if trust in management)
(d33)	Uncontrollable cause	A cause that is perceived as being not manipulable (d2, >< d27, ><d32)	Incompetency (if deemed persistent), rain (without trust in management)
(d34)	Causal frame	The way (mostly unconscious) in which the individual mind conceptualizes a stable causal model of their perceived reality (d2)	How floods work
(d35)	Values	Guiding principles in people's life (d2, d3)	Individualism, altruism, ecologism embodied in opinions about flood management.
(d36)	Beliefs	Beliefs about what is true or not (d2, d5)	"Urbanism causes floods."
(d37)	Norms	Rules, either formal or informal, that prescribes people's behavior (d2, d3).	The flood directive, cultural habits related to the water system.

The added value of perceptual approaches is that it increases the effectiveness of management, communication, and training policies (Bradford et al., 2012; Buchecker et al., 2013; Slovic, 1987). The biases and the differences in perceptions are instructive to the interactions between citizens and decision-makers, but also between engineers, natural scientists, and economists (Green et al., 1991). Investigating different perceptions of a problem means overcoming individual biases and allowing a more accurate representation of the problem (e.g., Boholm, 2009). Values (d35), Beliefs (d36), and Norms

(d37) are concepts part of the Value-Belief-Norm (VBN) model that explain people's behavior concerning environmentalism, along with comparable causal attribution terminology (Stern, 2000; Stern et al., 1999). The VBN model has gain popularity in explaining people's behavior in relation to water resources and risks as well, reviewed in Roobavannan et al. (2018), and from which definitions d35 to d37 are borrowed.

2.4.4.2 The Hidden Benefits of Relativism in Hydrology

Regarding the science of hydrology, how would this phenomenological perspective on causality relate to the science of hydrology? "*Man is the measure of all things*" allows the coexistence of multiple causal views, i.e., relativism, subjectivism, or indeterminism. This is a very contrasting viewpoint to the belief in universal laws or the analytical approach of universal reasoning (section 2.3.2; 2.4.1). Relativism is therefore disliked in science most probably because it opens the way to the first enemy of causality, skepticism. Still, it has many hidden advantages.

First, it is possible to restore the reputation of relativism in science by recalling that it is one of the drivers of its evolution. Indeed, relativism prevents science from falling into a second enemy of causality, dogmatism, or strict belief in a single truth, which is the end of the road for scientific inquiry. To a certain extent, all scientific knowledge is relative to time (Sivapalan and Blöschl, 2017; section 2.4.3.1). As such, the coexistence of competing true beliefs (d5) is a necessary condition (d20) for the growth of science (Kuhn, 1962) through the progressive refutation of the currently believed hypotheses (Popper, 1959). That said, if dogmatism is one dead-end, relativism through the pessimism or disenchantment that it engenders might be another one. Like Descartes' methodical doubt (section 2.4.2) or Hume's moderate skepticism (section 2.4.3.1), doubt or uncertainties in hydrology should be considered as the starting point for scientific reflection and speculation, not as its end or conclusion (e.g., Sivapalan, 2009).

Accordingly, it is relevant to question the causes that influence our community psyche and block us in a pessimistic and disenchanted status quo. While referring to the causal attribution terminology, some important paper's titles can be reinterpreted to extract a behavioral insight about hydrological disenchantment. Is "*Moving beyond heterogeneity and process complexity"* (McDonnell et al., 2007) moving from uncontrollable, stable, and situational

causes (d33, d30, d29) to dispositional and controllable (d28, d32) ones on which hydrologists can act? Similarly, is "*The secret to 'doing better hydrological science'*" (Sivapalan, 2009) just about taking dispositions (d28, d32) to move beyond uncertainties perceived as uncontrollable, stable, and situational (d33, d30, d29)? Finally, is dilettantism and fragmentation in the science of hydrology a stable trait (d28, d30) that cannot be changed or an unstable one that requires making efforts (d31, d28) (Klemeš, 1986a)? Hence, a recipe for breaking dead-ends and advancing the science of hydrology, or individual research, is to internalize these causes into dispositions, instead of attributing them to a situation (d29) originated from choices made consciously or unconsciously in the past that have reached their limits at this time (d30, d33).

Beyond scientific progress and the mood of perceptions, a certain skepticism may be associated with the relative plurality of representations within hydrology. It is manifest in hydrological modeling: there is a plurality of models, modelers, equifinal parameters, and ultimately a range of outputs that pessimistically describe our uncertainty instead of the certainties that a scientist wants to provide (Sivapalan, 2009). However, the struggle to recognize uncertainties has been difficult in hydrology (see Pappenberger and Beven, 2006), but fruitful in many ways. On the one hand, uncertainty turns out to be a scientific fact, and the proper characterization of uncertainty is knowledge (d6) rather than a lack of knowledge to the extent that it accounts for both what-is and what-could-be (d16, d17) and allows making better decisions (d2, d27). Moreover, uncertainty could be seen as the characteristic traits (d8) of catchments that could be useful to the task of catchment classification and comparison (e.g., Wagener et al., 2007; Chapter 4).

On the other hand, just as in water resource management (section 2.4.4.1), the plurality of causal perceptions could be considered as a reservoir of ideo-diversity, which both enables a more precise vision of reality and guarantees the resilience and adaptability of a science. For example, the apparent relativism in published articles makes it possible to conduct reviews and meta-analyses to extract the causes of hydrological concerns (e.g., Srinivasan et al., 2012). Secondly, different modeling philosophies or beliefs may have different merits and be useful in a different context (Hrachowitz and Clark, 2017). Moreover, relativism offers the opportunity to reason by abduction (d14) by building better models by studying alternative structures as hypotheses (e.g., Clark et al., 2008, 2011; Fenicia et al., 2011; Gupta et al.,

2012) that arise from the plurality of what has been increasingly recognized over the years as perceptual models. Interestingly, abduction (d14) was introduced by the phenomenologist philosopher Peirce (section 2.3.3.2). Its explicit acknowledgment is only referred to in recent papers (e.g., Baker, 2017; Harman and Troch, 2014). Finally, imagination (d15) is all relative: the creative mind is about portraying an idea (d11) to another scale, such as Newton pictorially did with the apple to the astral bodies (Bronowski, 1956), or "*the faculty of compounding, transposing, augmenting, or diminishing the materials* [d7, d11] *afforded us by the senses and experience* [d13]" (Hume, 1777).

Hydrologists have also enriched their concepts and perceptions by borrowing from other disciplines. Thus, visions of ecohydrology (Eagleson, 2002), sociohydrology (Sivapalan et al., 2012), or those of causal inference methods (section 2.4.3.2; Chapter 6) have been developed. The case of sociohydrology is particularly compelling from a phenomenological point of view. On the one hand, sociohydrology recognizes that water resources and the water cycle affect the development of human societies, as a source of well-being or threats, and that, in turn, human beings influence water resources and the water cycle, a fortiori in the Anthropocene era (Montanari et al., 2013; Vogel et al., 2015). Perceiving and realizing that humans influence the water cycle may seem trivial to many. However, hydrology did not take these coevolutionary processes into account in its numerical models (Montanari et al., 2013), nor in its perceptual representations of the water cycle (Abbott et al., 2019), and its recent recognition is invoked as a new paradigm (Vogel et al., 2015). Thus, hydrology has long remained faithful to Aristotle and to a two-thousand-year-old definition of nature that depicts it as that which has its principle in itself, far from technology and human influence (section 2.3.3.1). Sociohydrology breaks with this tenacious umbilical cord but does more in terms of phenomenological considerations. In seeking to explain and model (d8) in a mechanistic way (d9) the bidirectional coupling that exists between human perceptions, behavior, and hydrological systems (d7) (Baldassarre et al., 2013; Terpstra, 2011), sociohydrology explicitly considers variables representative of social perceptions such as values (d35, d2, d3), beliefs (d36, d2, d5), and norms (d37, d2, d3) as causes (d1, d7).

Finally, phenomenology, by recalling that everything is interpreted through the prism of an individually or socially biased consciousness, brings science and its method out of the stern rails of deduction (d12) and induction (d13). It

is not just a philosophical issue. Such biases have provoked a crisis of scientific credibility and reproducibility that is mainly felt today and related to phenomenology (Saltelli and Funtowicz, 2017). Indeed, just as the causal perceptions of everyone, the results of scientific research can be cognitively biased due to individual or socio-psychological mechanisms (see Andréassian et al., 2010; Nuzzo, 2015; Pfister and Kirchner, 2017). In order to remedy its biases (d2, d5) and to pursue the function of objectivity (d6) in its idealistic search for universal laws (d3, d10) or its realistic description of facts (d8, d12, d13), phenomenology paradoxically proposes to science to assume its part of subjectivity and to make science not solely abstract (d11) but be human (Bronowski, 1956). This is in that way that truth shall not be only discovered but created (section 2.3.2). Along with Hume (1777):

"Indulge your passion for science, […]*, but let your science be human, and such as may have a direct reference to action and society. Abstruse thought and profound researches I prohibit, and will severely punish, by the pensive melancholy which they introduce, by the endless uncertainty in which they involve you, and by the cold reception which your pretended discoveries shall meet with, when communicated. Be a philosopher; but, amidst all your philosophy, be still a man".*

In concrete terms, it is to organize science as a society of individuals with respectful interactions among divergent opinions and needs from within its body and outside with other social groups (Bronowski, 1956). It is to develop some adaptative mechanisms to control individual and collective biases, sectorial dogmatism, or on the contrary, the explosion of creativity and the dilution of knowledge. Such adaptative mechanisms exist and are continuously discussed: the peer review system (Pfister and Kirchner, 2017), community platforms (Ceola et al., 2015), transdisciplinarity (Montanari et al., 2013), sensing societal needs and promote them with community initiative and leadership (Sivapalan and Blöschl, 2017), the dissemination sciences into society (Kirchner, 2017; Lutz et al., 2018), the education of the next generation of hydrologist (see Seibert et al., 2013). Thus, a science or scientific community is an interacting and externally open system on an equal footing with a watershed or water management system. But then, what is a system? This is the last step of this evolutionary journey.

2.4.5 Systemic Thinking: Hydrology as a Whole

At the end of the 19th and early 20th century, the Newtonian mechanism and Cartesian reductionism were challenged from different scopes. In addition to the concerns discussed in section 2.4.2, Poincaré (1908), in "*science and method*", pointed out that Laplace's vision of randomness, as a lack of knowledge, is insufficient (d21) and that randomness may arise from sensitive dependence to initial conditions (i.e., nonlinearity; see section 1.1.2.3, Figure 1-5) making the world chaotic and unpredictable. He also discussed the issue of time reversibility in Newtonian mechanics, which was philosophically problematic in terms of causality regarding the principle of priority (d22) and physically, regarding the irreversibility of thermodynamical processes. Similarly, the Darwinian theory of evolution suggests that motion characterizes life (d9), but the Newtonian mechanics were inadequate to explain the emergence of complex forms of life and functions such as consciousness or voluntary motion. The philosopher Bergson (1858-1941 A.C.) insisted on the irreversibility of time for living organisms and regarding evolution. He has resurrected the final cause (d10) of Aristotle (Bergson, 1907). At that time, vitalism appeared as an anti-mechanistic philosophical trend that asserts that living things rely on independent principles that are non-physical and with unidirectional time. This is part of a more general philosophical trend, holism (Smuts, 1926), suggesting studying a phenomenon as an indivisible entity (d8). At the societal level, science became sectorized into goal-directed disciplines (d3, d10, section 2.4.4), heavily specialized and fragmented in such a way that you cannot understand what your colleague is doing in the next office or next floor because of the lack of unifying terminology in science, and the lack of universal scientific method (Poincaré, 1908; Wiener, 1948).

In this context, three influential trends have made their appearance to offer a unified theory and terminology involving system thinking and self-regulated systems: cybernetics, the general theory of system, and the theory of nonlinear dynamical system (or chaos theory). Whatsoever the trend, systemic thinking emphasizes the greater importance of the whole (d8) and its purpose (d10) over the parts (d7). The notion of a system as a whole, eventually made of interrelated parts but not necessarily, spread like a deluge in hydrology (Vemuri and Vemuri, 1970) and all the disciplines of knowledge (see von Bertalanffy, 1968). However, the notion of systems took different philosophical paths, with different meanings. Hence, a system is most likely

perceived differently (d2, d34) for hydrologists having a different background (section 2.4.4).

The first trend, i.e., cybernetics (Wiener, 1948), is inspired by control engineers, their guiding systems, and steering engines. Wiener and his colleague Rosenblueth realized that the term "feedback" is an essential characteristic of all goal-directed motion. They acknowledged the term teleology (d3, d10) as "*synonymous with purpose controlled by feedback*" in a taxonomy classifying the behavior of natural events (Rosenblueth et al., 1943). They build up their theory of systems on the recent advances made in thermodynamics, statistical mechanics, and their own development of a theory of information that had appeared synchronously in the work of Shannon (1948), allowing to account for the irreversibility of time and recognizing Bergson's philosophical argument. They give it the name cybernetics from the ancient Greek (*kubernetes*), meaning steersman, the same etymological origin of "governor", and illustrated the broad applicability concept on several topics. This trend is deeply rooted in system engineering and problem-solving. The linear theory of hydrologic systems of Dooge (1973), as an engineer himself, is imbued with Wiener engineering philosophy (see also Dooge, 1968) and contains some references to Wiener's work. To Dooge, as for other hydrologists with an engineering background (e.g., Chow et al., 1988), the system approach is an empirical black-box approach, an engineering term meaning that what is inside the box does not matter, convenient to escape hydrological complexity (section 2.4.2, 2.4.3; see also Chapter 1, section 1.1.1). The relationship between the inputs and the output is what characterizes the system (d8), and what allows "system operation" and "control", i.e., applied or operational hydrology (d10, d27). However, Dooge was rather a Newtonian mechanicist and determinicist hydrologist. The added values of the information theory, its analogy with thermodynamics, and pertinence to deal with thermodynamically irreversible processes in hydrology is now considered in the emerging debates about causality in hydrology (Goodwell et al., 2020; Koutsoyiannis, 2019).

The second trend is the organismic view in the general system theory of the biologist von Bertalanffy (1968). His book included the advances of his work that had started in 1920. It extends to a broader scope of Wiener's cybernetics as well as many other system approaches, although not convinced by the information theory. He stressed the critical characteristics of an open system, that is, to defy the law of thermodynamics by being organized thanks to

dissipative mechanisms of exchange of matter and energy with its environment (see section 1.1.2.4). The Aristotelean origin of his theory is entirely acknowledged (section 2.3.3.1). Besides, Bertalanffy also attributed to him the principle of emergence that "*the whole* [d8] *is more than the parts* [d7]". He is not a vitalist and believes that emergence could possibly be explained by the parts, but that is not the most effective way to address knowledge, that should be therefore rethought. Especially in biology, understanding an organ without considering its purpose or function is meaningless. In that regard, he also reconsidered the final cause (d10) and suggested five types of teleological behavior in natural events. The notion of equifinality now familiar in hydrology was borrowed by Beven from Bertalanffy (Beven, 2006a), and by transitivity, from Aristotle.

Table 2-5: Teleology (or finality) in natural events adapted from von Bertalanffy (1968)

Label	Term	Definition	Example
(d38)	Static teleology	A static arrangement that seems to be useful for a certain "purpose."	Optimality principle in ecohydrological distribution of plant species in watersheds (Eagleson, 2002); slopes directed to the outlet.
(d39)	Asymptotic teleology	Asymptotic behavior that attains a time independent condition.	A storm basin depletion.
(d40)	Emergent teleology	A purpose that arises from/in the organized structure.	Watershed's function as a support for life, ecosystems, and human systems (Kumar, 2007); physical habitat, food supply (Sivapalan, 2006); water partition, storage, release (Wagener et al., 2007).
(d41)	Equifinal teleology	Final state that can be reached from different initial conditions.	Beven's equifinality of model space (Beven, 2006a); a steady baseflow; equifinal flow paths to the outlet.
(d42)	True teleology	A true anticipated goal.	Water laws, directives, or acts;People's behaviors (section 2.4.4.1)

Besides, Bertalanffy stressed the importance of isomorphism (same form, d8) in science and nature, such that a unifying theory of science, or any discipline, should look for similar patterns in nature, in the scientific methods, or the scientific models. In this respect, isomorphism is, to some extent, the path recommended by the hydrological watershed classification (Wagener et al., 2007) or dimensionality reduction (see section 1.1.2.2). This recent trend in hydrology (see section 2.1) presents some analogy with the rise of the general

system theory considering the fragmented status of hydrology, the lack of unified terminology and knowledge of watershed processes at the macroscale, the non-parsimonious nature of physical reductionism, the previous calls for holism and synthesis (Blöschl, 2006; Sivapalan, 2003) as well as seeing catchment as a self-organized structure, like an organism (Kumar, 2007; Sivapalan, 2006). Also, Sivapalan (2006) stresses the interdependent importance of patterns (d8), processes (d9), and functions (d10), (d40), with the idea that function could be used to constraint patterns and processes. The latter reminds the importance of purpose-directedness and Aristotle's definition of randomness as out-of-purpose noise (section 2.3.3.1), but also its purpose-relativity (section 2.4.4). In general, the organismic view has spread since the '90s and is now considered as a new blueprint of the functional modeling approach relying on thermodynamics (Savenije and Hrachowitz, 2017). Despite the analogies, the current trend in hydrology is not an extension of that launched by von Bertalanffy's general theory of systems but often related to a Darwinian worldview by opposition to the traditional Newtonian worldview. The term originated from Harte (2002) and was merely defined. Harman and Troch (2014) redefined it as an approach that investigates the patterns of variation (d8) in populations of hydrologic systems and develops theories that explain their mechanism of emergence and coevolution (d9). Hence, Darwinian hydrology is not a functional systemic approach to modeling, and catchment classification is only a part of it. Since physical processes representation is eventually integrated into catchment classification frameworks, Darwinian and Newtonian hydrology could be pursued jointly (Sivapalan, 2018).

The third trend is built upon the deterministic chaos introduced by Poincaré (1908) and the development of the theory of nonlinear dynamical systems. It was made popular when rediscovered by Lorenz (1963) with computer simulations and termed chaos theory, emblematically represented by its butterfly. Although grounded in Newtonian mechanics, the theoretical development focuses on the property of the whole (d8) system by analyzing its attractor. The attractor is, such as Lorenz's butterfly (see Figure 1-4), the geometry drawn in the state space (d16) by observing the long-term dynamics of the system (d8, d10, d17, d39). This form is characterized in terms of geometrical complexity (or dimension) or in terms of sensitivity to initial conditions. The theory offers the opportunity of reconstructing a system from one single time-series (e.g., Packard et al., 1980; Takens, 1981) and infers its complexity and nonlinear behavior as a characteristic trait, namely a chaotic

invariant. The theory has been widely applied in hydrology (see Sivakumar, 2017) and is suggested for catchment classification (Sivakumar and Singh, 2012). In particular, some causal inference method is also able to reveal weak bi-variate causal interaction based on convergent patterns of isomorphism (d8, d24) between reconstructed attractors (Sugihara et al., 2012). An extension of the method allows addressing the principle of priority (d22) (Ye et al., 2015). Despite the challenges in applying the theory to hydrological data (Koutsoyiannis, 2006b; Sivakumar, 2000), the theory of nonlinear dynamical systems is congruent with systemic thinking and relevant according to the progressive recognition of the ubiquity of nonlinearities in catchment hydrology (Blöschl and Zehe, 2005; Delforge et al., 2020; Zehe and Sivapalan, 2009; see also section 1.1.2.3).

2.5 Connecting the Dots: Framing Causality

The evolutionary perspective of causality depicts a concept that was already plural and controversial from the starting point chosen in ancient Greece (section 2.3). Causality has continued to evolve and branch out over time, disconnecting and reconnecting concepts related to causality. Forty-two causal concepts were indexed and defined (regrouped in Appendix III). As noted in the review, plurality is a source of doubt, and one cannot portray the great diversity of causal concepts and leave the room while leaving it lying on the floor in pieces. This attempt at a conclusion where humanity has not yet come to a definite conclusion is, therefore, a first-person opinion on the subject.

Causality was introduced by Aristotle as an individual reasoning framework, aiming to answer “why” questions, i.e., producing the best explanations through four sub-questions: “from what”, “into what”, “how”, and “what for”. Causality was then progressively reduced to a body of mechanistic laws focused on the “how” that govern the real world and could be discovered by logic (section 2.4.2). We stopped calling these laws "causal" because of the blows to causality by Ockham (section 2.4.1), Hume (section 2.4.3), or Russel (1912), and we now call them physical laws, to guard against the tarnished reputation of causality. Successively, causality became an abstraction, a belief, a perception that is individually or societally constructed (section 2.4.4), which is a conflicting definition with a scientific mission that is supposed to provide objective facts and answers (Kirchner, 2017). Nevertheless, causality has never disappeared from our everyday language

and came back notably through systemic thinking, reviving the individual reasoning framework proposed by Aristotle (section 2.4.5), then in all scientific domains relying on different philosophical positions about causality (Figure 2-1, section 2.4.2 to 2.4.5).

The opinion expects to draw up a holistic vision of causality that allows it to retain its usefulness and universal character while escaping the highlighted epistemological issues. I conceive causality not as a body of established laws, not as an analytical framework, but as a system coupled between nature and society that aims to produce not ultimately the best but better societally acknowledged explanations or knowledge (or the best at a time-point of history). Indeed, the distinction is essential as it replaces causality in the temporal and evolving frame that has been emphasized, allowing both continuous progress in science and in our understanding of causality. A conceptual model in Figure 2-3 illustrates this system.

The model includes a divine world and a world of ideas to remind that causality was also historically attached to these worlds through providential (section 2.3.1) and transcendental (section 2.3.2, Figure 2-2) causation and that such views of causality remain present in society. My intention is not to judge anyone's personal beliefs but to point out that these beliefs exist, making them perceptions, real within the real world, that a scientist may need to account for through a phenomenological approach (section 2.4.4). As a scientist, the system of interest is specifically causality in the real world. The real world encompasses in a fuzzy way Particulars, i.e., objects that can be observed (nature, human, society, or coupled system), the individual himself and his/her mind, and society. Thus, I gave up on the sharp distinction between human and nature since they became less and less separable through time, given the human inability to escape himself while observing and its increasing influence on nature and its behavior. Within the real world, I deliberately chose to place the human mind at the center of the model because everyone approaches causality through their individual experience. However, this human mind could be seen as well as a collective mind representing a group of people that shares the same causal frames and explanations.

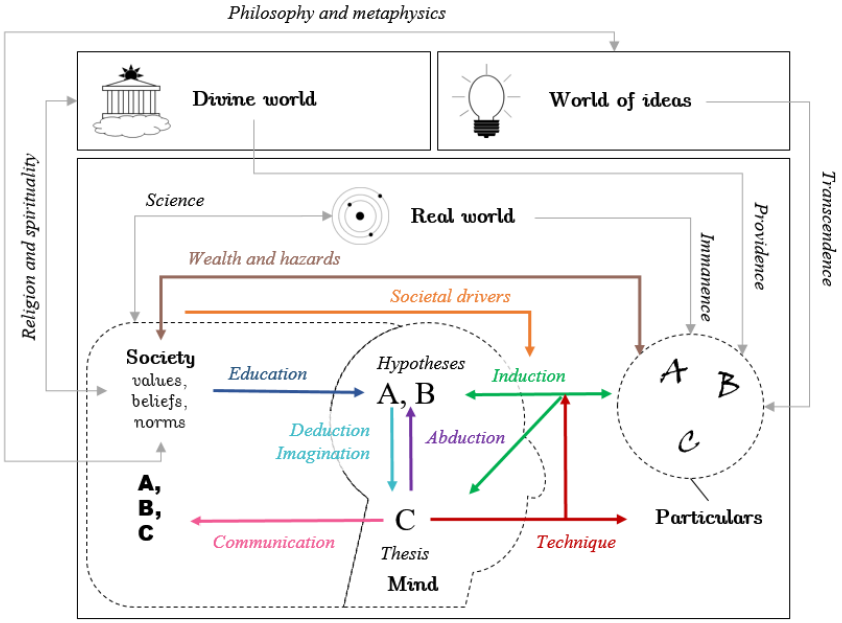

Figure 2-3: A conceptual and dynamic model of causality as a system producing explanations. Explanations can be grounded in different potential substrates: the divine world ruling the real world by providence, the world of ideas, in reference to Plato, affecting the real world by transcendence, and the real world ruled by itself (immanence). Within the real world, three other substrates are considered: particulars as concrete and observable objects, a mind of an individual or collective agent, and a society defined by its values, beliefs, and norms. Particulars A, B, and C become abstract objects or hypotheses in mind, where A and B explain another one C. The full abstract representation is constructed by the mind following three forms of causal inference or logic (induction, deduction, and abduction), or with flexibility using imagination. Induction is bidirectional because not neutral and full of hypotheses that should be discussed. Abstractions could also be socially inherited from others through education and learning. Technique refers to the processes of intervening on particulars (engineering or arts), with the eventuality of reinforcing induction based on the technology of observations, analytical methods, and hypotheses testing. Communication refers to the act of sharing abstractions with the society to modify or reinforce the socially constructed abstractions of A, B, and C. Societal drivers can influence induction as they provide focus and intentions to the individual minds, as well as a structural framework for observation and its financing. Particulars representing all investigable objects (in nature or human systems) may be affected by or affect society by a bidirectional arrow of wealth and hazards, affecting our interest in understanding the real world.

Within the mind, explanations are produced to interpret and structure Particulars in the real world, to act on them, and to share these explanations with others in society. Explanations can be seen as an organization between thesis and hypothesis that structure reality into a causal frame (d34). The internal mechanism of this causal system producing explanations involves the three types of logic plus imagination to produce and reframe abstractions that serve as an explanation (Table 2-2, or Appendix III, d11 to d15). Observation through induction could be driven by the Aristotelean framework of causation with whatsoever focus on one or several causes among the four. Induction is also influenced by cognitive biases such as antecedent knowledge, peoples individual purpose, or societal drivers that encourage people to investigate and discuss particular purposes, e.g., on topics that concern wealth and hazards interactions between society and particular objects, values, beliefs, and norms (Table 2-4, or Appendix III, d35 to d37). To overcome bias, acting on the real world to test, verify or falsify, the suitability, effectiveness, or manipulability of our causal frames (*Technique* in Figure 2-3) and refine induction is the crucial driver to reframe and improve explanations. Another driver is exchanging with society to share our explanations and enquire about possibly better ones.

Now that this systemic view on causality is depicted, we see its elements, its organization, mechanisms, and its end of production better explanations. As carrot and stick, causality is both seen as a purpose, an ideal, the goal of science, or an end in itself (understanding), and a constrained and organized framework relating specific and distinct approaches, philosophy, and tools together. From a temporal perspective, causality is threefold: a capital of knowledge accumulated and inherited from the past, reinvested as a day-to-day tool, for its future growth. The opinion can be clarified and refined through answers to exciting questions related to epistemological issues highlighted in this review.

1. *What are a causal explanation and its difference from other explanations?*

The purpose of the causal system is to propose the best explanation possible in our historical time. So far, the best explanation could be the holistic one that has most of the traits related to causality (d43, see also Appendix III).

(d43) Causal explanation: (1) a stable agreement, perception or belief emerging from a collective thinking constrained by logic in interaction with the real world [d4, d5, d6, d25, d28, d30, d34, d36, d38-42] … (2) explaining within the collective context and for a specific purpose [d2, d3, d10, d29, d35, d37, d39, d42] the what-is-to-be [d8, d11, d16, d38-d42] through mechanisms [d1, d9, d19, d24] linked by inference to other elements [d7, d12-15], either in space [d23], in time [d22], or both and possibly at other scales, in an intelligible but sufficient way [d18, d21, d26] … (3) and by virtue of a potential future practical application [d2, d3, d10, d17, d27, d32, d41] enabling either testability, technological progress or successful control while operating and intervening on the real world.

This holistic definition reflects a community objective. As a result, individual scientists partially contribute to the production of a causal explanation through their day-to-day objectives. Any partial explanations could be seen as non-causal but potentially contributing to a causal explanation (e.g., a theory that is not tested or not acknowledged by the community, a statistical association without an account on the mechanism). There is a continuous range of partiality between a partial and a causal explanation, allowing to say that an explanation is more causal than another one. Causality is thus relative, possibly absolute, through its unanimous acceptance. However, history showed us that many scientific propositions judged absolute could become wrong or relative (e.g., Newton's Law of Universal Gravitation).

The definition (d43) allows giving spaces to different ways of doing hydrology. The first part (1) stresses the importance of the societal context and portrays the states of causality as a consensus. Hence, in doing so, it is possible to do hydrology by sensing people's perceptions of water-related issues for water management (section 2.4.4.1). It also encourages individual scientists to deal and strengthen themselves by dealing with the present plural causal views 2.4.4.2), for instance, by doing reviews, meta-analyses or

syntheses (e.g., Blöschl, 2013; Srinivasan et al., 2012), or by working with multiple hypotheses (Beven, 2012a; Clark et al., 2011). The second part (2) portrays the facets of causality following the Aristotelean framework and its four causes (section 2.3.3.1). Again, individual hydrologists could invest themselves flexibly into such a framework, studying preferentially mechanism, lower representative elements, their interrelations, or the whole system and its patterns. These are common ways to understand catchment systems within the science of hydrology. The third part (3) implies that theories or models should be applicable so that our understanding could be tested, but also for the practical purpose of operational hydrology.

2. *Are causal explanations real or abstract?*

Causal explanations are abstract by being the product of reasoning (d11), such as a hydrological model, a theory, a law. In this sense and at the individual level, causality is not discovered but created (Bronwowski, 1956) by the mechanism of Figure 2-3. Hence, causality is not reality, and the causal explanations, laws, or models do not have any substance in the real world or any metaphysical world. This view of the external laws of causality was rejected by Russel (1912). If causality does not exist in the real world, it is grounded in our minds while not being neither a pure product of our fantasy (d15). The connection between reality and our minds are our lived experiences where our thoughts interact with the real world (d13). Hence, causality is undoubtedly an abstraction. However, if not real, it is allowed to be realistic since reality is embedded in causality thanks to our interactions with the real world (similarly to Figure 1-4). One could infer some patterns related to geomorphology, rainfall, evapotranspiration from the hydrograph (e.g., Kirchner, 2009). Streamflow is not a catchment, rainfall, neither evapotranspiration. Still, approaching one from another is possible because constrained processes bound these hydrological elements. In the same way, causality is an output of the mind's logically constrained interactions (d11 to d14) with reality and thus captures and channels something about reality. Such pieces of reality do not have to take the form of a physical realism (see Beven, 2002a); they are realistic for their capability to be justified and acknowledged as true (d6).

It is essential to recognize that causality is linked to reality (while not being reality) as this cuts short a Manichean debate on the nature of models or theories. Yes, all models are wrong; if wrong means abstract in their substance

(section 2.4.1). While this is a good reminder for those who confuse reality with abstractions of it, it is not a sufficient reason to disengage scientists or tarnish their motivations in their quest for realism or truth because it relies on a necessary third concept, causality, that allows models to be right for the right reasons (Kirchner, 2006).

3. *Are causal explanations true or consensual?*

As mentioned, a causal explanation can be true in the sense of realistic (constrained by logic) and operational. However, a trait of a reasonable and logical explanation, a fortiori a causal explanation, is still to be understandable (d18) and accepted. The elements related to perceptions, education, and communication are not to be neglected in the question of causality and, therefore, in hydrology (e.g., Bogaard et al., 2017; Lutz et al., 2018; Seibert et al., 2013). The question is: can we speak of a causal theory, model, or explanation if it is not acknowledged? Given the proposed foundations for causality, the universal trait of causality is no longer linked to its universality in the real world but rather to the consensus on the broad applicability and explanatory power that a way of understanding has acquired within a community. Best scientific explanations have to fully meet the community to fully deserve their causal status, and an outstanding paper will fail to attach itself to causality if it remains unseen. Then, causality is also the product of a social system that also constrains, together with our experiences of particulars, what may or may not be judged right to achieve a societal construction of causality around a consensus. Sometimes, it is done wrong, as evidenced by the Galileo trial and the past vision of geocentrism. Accordingly, a view of causality only as a social and subjective construction of reality is dangerous and should be balanced with already acknowledged causal explanations, experiences, data, and facts. However, hopefully, social interactions and dialogues allow correcting individual scientific biases through the sharing and mutualization of experience for the sake of truth. Causality should be realistic and consensual. With an increasing level of consensus and adequacy with reality, we can progressively strengthen the terms causal explanations to causal theories, paradigms, or laws.

At some point in history, a consensus on an explanation may not be reached due to a lack of knowledge or philosophical contradictions. Hopefully, the causal system has some memory and transcends time by storing and releasing explanations later on as a hydrological reservoir. One example of this was the

Aristotelian conception of causality that resurfaced in a general theory of the system or in a science of hydrology that is relatively more focused on the questions of representative elements, patterns, and function (section 2.4.5).

4. *What is the relation between Causality, Complexity, and Uncertainty?*

Rationalism (section 2.4.2) and the principle of sufficient reason (d19) has plunged humanity into a deterministic vision of causality, such that causality was opposed to uncertainty or randomness (e.g., Koutsoyiannis, 2010). However, if uncertainty is recognized as a fact and a subject of study (Pappenberger and Beven, 2006), causality applies to it, and uncertainty is no longer in direct opposition to causality. One can ask oneself about the causes of uncertainties: their source (d7), their form (d8), their propagations (d9), and their meaning and purposes (d10) (see also Nearing et al., 2016).

Hydrology deals with complex systems. If complexity could be described in terms of intrinsic properties of systems as nonlinearity, dimension, or lack of organization (see sections 1.1.2.2 to 1.1.2.4), another feature and trait of complexity is our difficulty to understand resulting in plural and coexisting causal views or models (see section 1.1.2.1, Figure 1-2). Epistemic uncertainties are inevitable while dealing with complex systems (e.g., hydrological, social, or climate system) and will most likely never disappear. Their descriptions are beneficial because better models and predictions can be achieved with modeling frameworks (e.g., Clark et al., 2008, 2011; Fenicia et al., 2011; Gupta et al., 2012) or only better prediction through an ensemble model.

However, plural causal views should not explode, and it is the duty of science to limit uncertainties as much as possible, improve data and logic to constraint their extent. For that purpose, it appears that scientific disciplines dealing with complex systems show a stronger affinity with the concept of causality (Figure 2-1). In contrast, physics that deals with simple systems and well-controlled experiments (e.g., a pendulum) is more easily freed from the concept of causality, talking about certainties, and focuses on acknowledged mechanisms (d9). Still, on the scale of centuries, the great revolutions of physics are similar to changes in systems of reference (Kuhn, 1962), i.e., a reframing of formal causes (d8), which, in result, offers a better representation of mechanisms.

5. *Does Science Make Progress?*

If portraying uncertainties is, in fact, science, we are entitled to wonder if science is making progress? A distinction must be made between technical progress and progress in terms of understanding. Technical progress is manifest and measurable within society. However, progress in terms of understanding is more difficult to establish. To a certain extent, technical progress may incorporate advances in our understanding, but it also incorporates social values, demand and financial capital so that new technologies do not always reflect progress in terms of scientific understanding. Similarly, we cannot assume either a positive or negative correlation between progress in understanding by counting scientific communications.

Today, the way we portray the growth of our understanding is still idealistically represented by a growing trend, punctuated by cycles of enchantment and disenchantment (Kuhn, 1962; Sivapalan and Blöschl, 2017). Until now, does understanding grows, or is it a tacit hypothesis? The idealistic growth somehow relies on a transcendental view of causation (section 2.3.2, 2.4.1, Figure 2-2), such that understanding is a one-way transfer of causality from either the world of ideas or the real world to the individual minds. The former is questionable and not scientific. The latter is an individual process through induction (Figure 2-3), a vision of personal understanding, and thus not representative of the growth of understanding within broader groups of people such as society or a scientific community. In fact, the only way one can really monitor the growth of our understanding is by measuring and tracking for a given topic the level of interest and consensus on that topic (*Education* in Figure 2-3, e.g., by doing scientific reviews, bibliometric analysis, text mining). In that way, we can actually check if the assumed growth is verified and portray cycles of enchantment and disenchantment as periods of convergence and divergence of opinions.

This evolutive perspective on causality suggests that progress in understanding is not that linear but subject to complex patterns. Similarly to recent visions of the evolution of species, Ideas compete and co-evolve, branch and hybridize, gain and lose importance, die, resurrect or explode, such as the ideas on causality presented in this review. Some will succeed in becoming static by being accepted forever by humanity or a scientific community. This is the fulfillment of their causal status.

6. *Can we rely on a reduced vision of causality?*

Before synthesizing a holistic view of causality, this review showed that particular philosophies of causality, whether rational, empirical, phenomenological, or systemic, have always enabled the development of valuable knowledge, approaches, methods, skills, or technologies. Pursuing distinct or reduced definitions of causality is a source of progress by contributing to the ideo-diversity and the pursuit of causality from different fronts. Each scientist or scientific discipline needs to focus on specific objectives, boundaries and limits, a manageable amount of methods, and a specific vision of causality to expect to produce an organized scientific output. In doing so, both scientists and scientific disciplines are limited. Hydrology will never say everything about water, just as biology will never say everything about the specificity of living things. Overcoming the limits for the sake of a better understanding and management of complex systems is essential and a matter of transdisciplinarity (e.g., Abbott et al., 2019; Lall, 2014; Montanari et al., 2013; Savenije et al., 2014; Vogel et al., 2015; Wagener et al., 2010).

As long as a science is not in a pathological case of fragmentation, it is then perfectly acceptable if not recommended for a scientist or a scientific community to remain focused on defined trajectories and causal preconceptions because this opinion supports that causality is solved a the societal level, and only partially solved by lower levels.

7. *Thinking causality: what for?*

If anybody can pursue one's approach to causality, why invest in extensive reflections on causality? Figuratively speaking, by borrowing Sivapalan's metaphor (2006), scientists and scientific disciplines are like musicians and ensembles within an orchestra. If plural and reduced causality projections may present a risk of cacophony and some pessimistic sense of fragmentation, thinking causality is studying harmony and music theory, a common language for musicians allowing them to play and discuss together. This common language was the dream of the first theories of systems (section 2.4.5; Bertalanffy, 1968; Wiener, 1948) and explicitly focused on causality at the origin.

From an operational perspective, thinking causality means adhering to the fact

that a causal view offers better than any other view the possibility of managing complex systems, especially those changing (section 2.3.3.3), with some chance of success. With the increased difficulty in the tasks related to hydrology in the Anthropocene era (e.g., Clifford, 2002; Montanari et al., 2013), failure can significantly impact people's lives or the environment. To be right for the right reason (Kirchner, 2006) is a duty or responsibility that is difficult to achieve but one that we must strive for at different levels, as a scientist, a community, or a society.

Importantly, through a social and logical view of causality, which always aims at success, the sources of failure may arise not from an error of logic as such but from a neglect of social values, beliefs, norms, or perceptions (unless ignoring them is an error of logic in itself). Succeeding is indeed just not about reasoning the causes and effects (d1), but debating them (d3) to harmonize perceptions for a consensus allowing the success of collective action (d2). This is the point behind participatory and integrated approaches in hydrology. From a phenomenological and social perspective, thinking causality is improving your ability to connect with people through exchanges of causal frames. It boosts our ability to be understood by adopting a causal language that activates the causal frames of others, and our capability to understand.

Indeed, at the scale of reasoning, inquiring about how other scientific disciplines invest causality is a profound source of inspiration and plasticity for scientists' minds. Thinking causality gives them open-mindedness, resilience, the ability to find innovative solutions and escape dead-ends by seeing what needs to be done (see d28 to d33). Investigating several approaches, alone or with others, in parallel or in series, is a necessity. For example, the literature review showed that physically-based models would benefit from being further simplified and from testing the assumptions on which they are based. This route is taken notably through empirical methods and an empirical philosophy of causality. Conversely, empirical methods benefit from being constrained by theoretical principles and from being linked to mechanisms. Whatever the approaches, they are complementary and do not free themselves from each other. They are destined to converge so that between competing theories, perceptions, data, and methods, everything converges towards a theory minimizing uncertainty, and that can be called causal.

On a broad scope, our vision of causality should arguably map to our vision of the scientific method given their shared purpose of making better explanations. The latter was mainly constructed by confronting the logic of deduction with that of induction (see Figure 2-3). This scientific method was already a synthesis of rationalistic and empirical visions of causality (sections 2.4.2 and 2.4.3), a view still commonly used today (e.g., Box, 1976; Dooge, 1986). However, such a prevailing model dating from the 18th century leaves little room for the importance of community and society in the method. It fails to include the logic of abduction (d14) allowing to infer from multiple hypotheses (Baker, 2017; Harman and Troch, 2014; Peirce, 1960) and does not account for the importance or even dominance of societal drivers in our conceptualization of the scientific production process (analogous to Abbott et al., 2019). Still, Dooge (1986) had concluded about the finding of hydrological laws: "*Hydrology can establish itself as a science but not without a degree of organization* [d40] *in planning and in thinking that has not been evident before now*".

Interacting with the scientific community or society, communication and teaching skills, making synthesis and reviews, and taking scientific leadership are not soft skills for scientists. All processes in Figure 2-3 appear on an equal footing. Of course, no one can be involved in all of these tasks, especially not at the same time. Causality is once again the concern of a community and a society. Even within a society, there are times for everything, e.g., to be more synthetic (Blöschl, 2006), to give focus (e.g., Blöschl et al., 2019; Sivapalan, 2003), to be more rational in our science or communication (Kirchner, 2006, 2017), to be more empirical (Sivapalan, 2018), or to collect data (Beven et al., 2020). Indications of what we could do are given in Figure 2-3. We should not do "more" of everything all the time. What we should do as a community depends on the state of the causal system so that it can be balanced and maintain its mission of providing better answers to society without falling into pathological cases.

2.6 Conclusion

Our perception of causality keeps evolving, but through the perspective brought by its evolutionary journey, constants remain. The interest in causality, whether explicit or silent, sustains as a human necessity to understand. In ancient Greece, causality was already perceived as a unifying ideal to pursue, a framing of how things work that is robustly assessed by science and logic, or an agreement achieved within society by communication to take actions. It is still valid while looking at the specific concerns of hydrology (section 2.1). In particular, causality was explicitly formalized by Aristotle (Falcon, 2019), together with the logic of deduction and induction that is the basis of a scientific method rolling on theorization and experimentation.

Through time, causality was reduced to simpler formulations. In the 17th century, rationalists focused themselves on processes and deduction as hydrologists working on physically-based models. In the 18th century, empiricists focused on generalization from observations by induction, as well as empirical modelers of hydrological systems. These two Manichean views of causality are inseparable in hydrology because models, even if physically-based, are relying on the inductive logic of calibration, and empirical models themselves make assumptions, if only on the representativeness of the data. The plurality of models has forced hydrologists to call them perceptual, recalling the phenomenological current appearing at the end of the 19th century, which reminds us that the object of our scientific investigations is the biased and limited perceptions of things and not the things themselves.

Perceptions are now ubiquitous in hydrology or in water governance based on the participatory process that sense perceptions and promote communication among stakeholders. Due to the problem of induction stating that wrong generalization can happen due to limited experimentation, all causal laws became beliefs or perceptions (Hume, 1738) that should be falsified (Popper, 1959) since experimentation is always limited in space or time. Nowadays, empirical methods regain popularity to infer causality and conceptual models from data (e.g., Goodwell et al., 2020). Besides, flexible physically-based modeling approaches are now comparing different perceptual models to infer the best hypothesis, which follows the logic of abduction formalized by the phenomenologist Peirce (1960). Sociohydrologists build human-water coupled systems based on people’s perceptions (Sivapalan et al., 2012).

Causality in science tends to a universal and robust way to harmonize and to deal with perceptions instead of revealing single truths as absolute laws. The theory of systems itself was introduced to harmonize a fragmented science around one reasoning framework that is closely related to Aristotelean notions of causality (von Bertalanffy, 1968) that stresses the importance of patterns, processes, and functions (similarly to Sivapalan, 2006).

Regardless of the route taken to investigate causality (physical, empirical, perceptual, or systemic), hydrology has evolved fruitfully by taking them separately but was, however, perceived as fragmented. The current trend to unify our understanding of the hydrological systems seems to connect these visions of causality and make them coevolve. In this regard, we propose a definition of causal explanations and synthesized model of explanations that expands the scientific method further than the sole logic of deduction and induction, to account for the social and systemic sides of causality. Causality is depicted as an evolving system providing causal explanations within society from the logically constrained interactions of our minds with the real world. The opinion suggests, therefore, to move away from a widely adopted preconception of causality related to deterministic and mechanistic laws. Instead, causal explanations are measured as an emerging, robust, and useful collective agreement on how systems are conceptualized, including through their relevant mechanisms, for a given context and purpose such as control.

Chapter 3 Study sites and data

"It is beyond a doubt that all our knowledge begins with experience."

Kant

Abstract

Next to the city of Rochefort, the Lhomme Karst System (LKS) consists of a limestone outcrop located on either side of the Lhomme river on the downstream part of its catchment. The Lhomme waters contribute continuously to the karstic aquifer's recharge before being returned to the river by a Vauclusian resurgence. An extensive network of cave and conduits developed in these limestones and one of the largest cave in the system has been invested for more than twenty years by a scientific research laboratory, i.e., the Rochefort Cave Laboratory (RCL). The thesis uses a few selected datasets to develop the case studies covered in the following chapters. Chapter 4 focuses on three discharge time-series obtained from gauging stations located before, inside, and after the LKS, to study karst induced nonlinearity. Chapters 5 and 6 use an Electrical Resistivity Tomography (ERT) dataset representative of the subsurface epikarst above the cave at RCL. The ERT dataset is investigated in order to find spatial patterns based on dynamic similarity. Besides, Chapter 6 includes other environmental time-series: relative gravimetry monitoring mass changes in the system, evapotranspiration and rainfall data, percolation drip discharge data monitored from within the cave, and groundwater level data showing high peaks whenever the river overflows and floods the cave system. These are used in addition to the ERT dataset to develop study cases on causal inference between hydrological time-series.

3.1 Lhomme Karst System (LKS)

The Lhomme (or Lomme) Karst System is located in the downstream part of the Lhomme river basin in southern Belgium (Figure 3-1). The karst is composed of limestone rocks dating from the Givetian period (Middle Devonian, around -385 Ma). The system is at an advanced karstification stage, with an extensive network of conduits and caves (Bonniver et al., 2013; Hallet and Meus, 2011; Michel et al., 2015; Poulain, 2017). The karstic outcrop is oriented in the same SW-NE axis as the river. The Lhomme River crosses the karst from the confluence between the Lhomme and its tributary, the Wamme river (Figure 3-1).

When the river leaves the karst area, the waters of the karst aquifer are discharged into it through a Vauclusian spring, which is the aquifer outlet (Appendix VII, Photo VII.1). Along this 7 km strip, the river is perched, which means that it is disconnected from the saturated zone and is no draining the aquifer. Instead, it is the river water that percolates through the karstic riverbed to recharge the Lhomme aquifer. As a result, the system contains two underground rivers that flow parallel to each other (see also Delforge et al., 2017). One is the underground Lhomme, and the other is the underground Wamme, as it starts upstream of the confluence from the section of the karstic system crossed by the Wamme at the surface. Then, the karst aquifer is permanently under the influence of recharge fluxes (I in Figure 1-1), thus not meeting, in addition to the hypothesis of aquifer homogeneity, the hypothesis of absence of recharge that would allow inferring the hydrogeological properties of the aquifer from the parameters of the B&N model (Eq. 1.5). To limit percolation and excessive depletion of the river, the river bed was historically paved. Episodically, quick recharge of the karst aquifer also occurs during flood events, whenever the discharge at the city of Rochefort (station S2, Figure 3-1) is above 15 m^3/s (Poulain, 2017). The karst system is, therefore, a natural storm basin (Appendix VII, Photo VII.2).

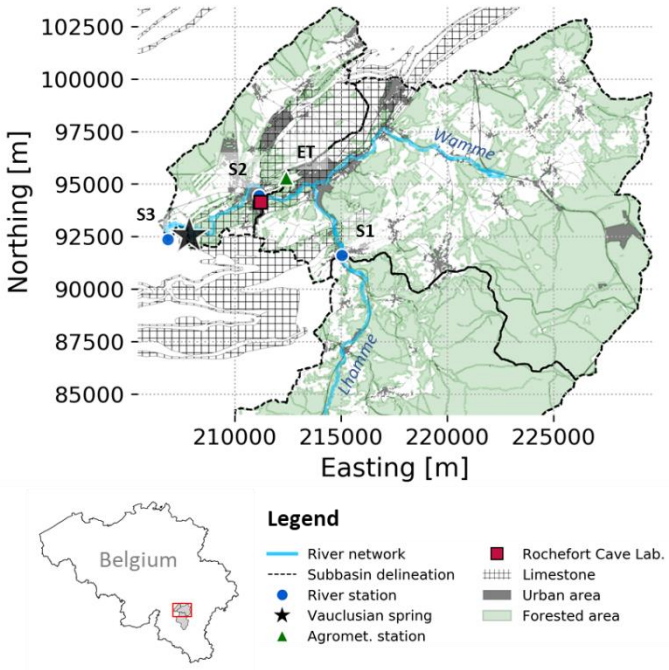

Figure 3-1: Lhomme Karst System (LKS) and the downstream part of the Lhomme River Basin (EPSG: 31370; data source: SPW). After its confluence with the Wamme at the city of Jemelle, the Lhomme flows through the LKS for ~7 km. Along this section, the river is perched, and the Lhomme contributes to the karstic aquifer's recharge. During flood events, the Lhomme overflows into the cave systems leading to a rapid recharge. The aquifer returns the water back to the river at the Eprave Vauclusian spring. S1, S2, and S3 are SPW river gauging stations on which the subbasins delineation is based (see section 3.3.1). Potential evapotranspiration data (ET, section 3.3.3) are from the PAMESEB agrometeorological station of Jemelle. The Rochefort Cave Laboratory (RCL) is installed in one of the major caves present in the LKS.

From the western side to the northwestern part that delimits the karst, the system is surrounded by Frasnian shales (Late Devonian, around - 377 Ma). This impervious boundary encloses the system and forces the water to resurge at the Vauclusian spring. Located on the banks of the Lhomme, the spring returns the groundwater to the river and is the only known outlet of the karst system. However, the groundwater discharge is problematic to gauge at the spring due to its proximity to the Lhomme river. The Lhomme water potentially intrudes into the spring area caused by its relatively high discharge and water level. The resulting impedance limits the immediate release of the spring water into the river. Poulain (2017) performed modeling and gauging experiments before and after the confluence of spring discharge into the

Lhomme. He reports an estimate of the average spring discharge of 1.1 m^3/s for the period 2013-2015 (i.e., 20% of the mean discharge monitored at S2, Figure 3-1) with an amplitude of 0.6 m^3/s to 3-5 m^3/s. On the 12th of October, 2018, the University of Namur conducted another gauging while the Lhomme river discharge was close to zero m^3/s due to an unusually dry summer (Appendix VII, Photo VII.3). They reported a spring discharge of 0.50 m^3/s, while the mean daily discharge at S3 (Figure 3-1) was 0.56 m^3/s. Hence, the Vauclusian spring is the primary source of drought resilience for the river system downstream.

3.2 Rochefort Cave Laboratory (RCL)

Standing 50 m above and located 300 m south of the Lhomme river (Figure 3-2), the Rochefort Cave Laboratory (RCL) monitors the Lorette cave from both the surface and the inside. The Lorette cave is one of the biggest explorable cave within the LKS, with a network of galleries extending over 500 m along the W-E axis and 200 m in the N-S axis. The eastern part is reserved for tourist activities while the central part (Figure 3-2, B) is devoted to scientific research since the 90s (Camelbeeck et al., 2012; Quinif et al., 1997) and was equipped with new instruments since 2013 after the launch of the KARAG project (see www.karag.be, and section 1.2.1). On the cave surface, a shelter houses a relative superconducting gravimeter (Goodkind, 1999), monitoring gravity changes (RG, Figure 3-2, B; Appendix VII, Photo VII.4, and VII.5). Local gravimeters are sensitive to change in the mass distribution above and below the ground. The high resolution of modern gravimeters offers the opportunity to close the hydrological mass balance (Eq. 1.1) and to monitor or model hydrological processes (Figure 1-1), such as precipitation, runoff, evapotranspiration, infiltration, groundwater recharge and depletion (e.g., Creutzfeldt et al., 2014; Delobbe et al., 2019; Hasan et al., 2006; Jacob et al., 2008; Van Camp et al., 2016; Watlet et al., 2020). In particular, the gravimeter captures the flood peaks that occur when the river overflows into the cave system, as well as the groundwater level data (GL in Figure 3-2.A) since the water directly flows into the cave where the sensor is installed (Appendix VII, Photo VIII.2).

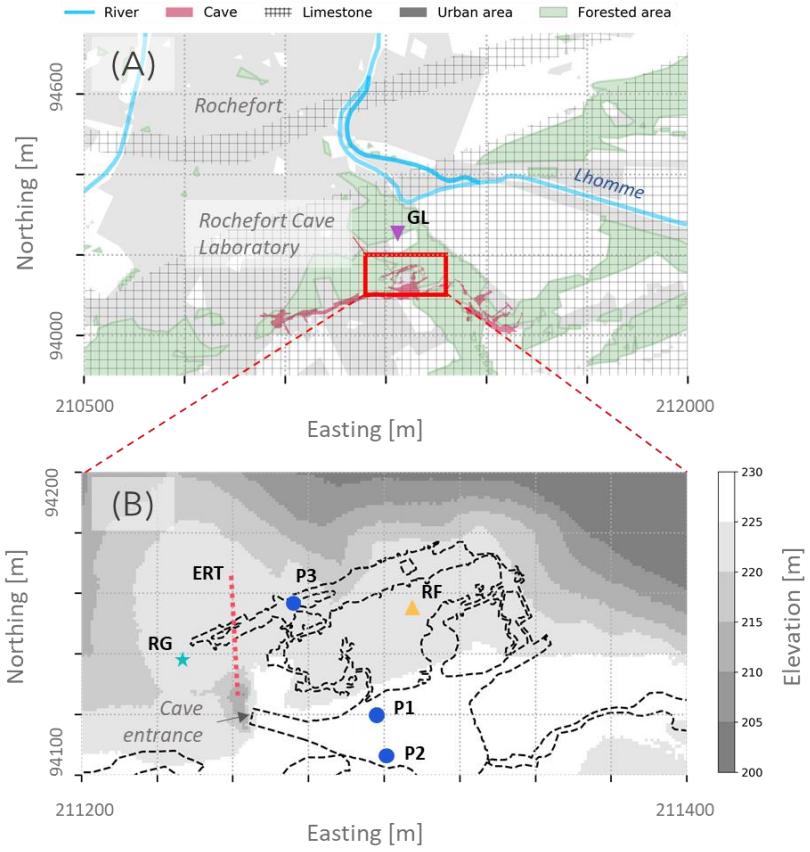

Figure 3-2: (A) Location of the Rochefort Cave Laboratory (RCL), and (B) its instrumental set-up specifically related to data used in the thesis (section 3.3.3, RG: surface relative gravimeter; ERT: Electrical Resistivity Tomography profile; P1 to P3: drip discharge percolation monitoring within the cave; RF: surface rain gauge). Groundwater level data (GL in A) are those of the cave of Nou Maulin water table. Coordinate system: EPSG 31370. Data source: SPW (river network, landcover, karst delineation, elevation), and Watlet (2017, cave delineation, and instrument coordinates).

Also, at the surface, a Lufft tipping bucket type rain gauge monitors rainfall, and an Electrical Resistivity Tomography (ERT) profile (RF, ERT, Figure 3-2, B) was installed to investigate the subsurface above the cave, also known as the epikarst. The profile is not flat because it starts from the depression of a sinkhole (Appendix VII, Photo VII.6), where the entrance to the cave is located. A staircase gives access to the cave entrance from the shelter to enter in the Val d'Enfer room where the two drip discharge monitoring devices (Kaufmann et al., 2016) are installed (P1, P2, Figure 3-2, B; Appendix VII, Photo VII.7). In particular, P1 monitors an active dripping point linked to the

presence of a visible fracture on the ceiling of the chamber. Based on dye injection at the surface and in-cave tracing, Poulain et al. (2018) identified a connection and preferential flow path between the dye injection point and P1. The breakthrough curve showed an initial arrival time of 3.75 hours, a sustained peak for 80 hours, and a tail lasting up to 120 days. These numbers are relative to a diffuse flow. However, sporadic peaks in concentration were observed after every rainfall event, reacting after 1.48 hours, peaking after 7.2 hours, and lasting up to 30 hours on average. P2 monitors a dripping spot draining a porous limestone area. The last one, P3, located in the North gallery, is a device monitoring slow discharge from drops falling from one single stalactite below a massive limestone layer. Within the KARAG project framework and the two associated theses (Poulain, 2017; Watlet, 2017), these datasets have already been valorized in several publications. They describe them in more detail (Poulain et al., 2018; Watlet et al., 2018b, 2020).

3.3 Thesis Datasets

A part of the MIGRADAKH project (section 1.2.1) was to gather and harmonize the available environmental data and metadata for the LKS. This task was carried out during the first two years of the project. Appendix II describes the resulting database and metadatabase. The case studies developed in the thesis are based on only a small part of the data available in the database. This section describes only these relevant subsamples of data.

3.3.1 Streamflow Time-Series

Daily streamflow time-series from 2004 up to 2010 (7 years) are obtained from hourly discharge data gathered from three gauging stations (S1, S2, and S3, Figure 3-1). These belong to the same monitoring network of the Public Services of Wallonia (SPW) and are publicly available from their data portal (aqualim.environnement.wallonie.be). The S1 station is located just before the karst and delineates an upstream basin of 247 km² covered by 70% of forested areas. The S2 station, delineating a 424 km² watershed, is located on the Lhomme karst system in the Rochefort urban area after the confluence with the Wamme river in the city of Jemelle. The third station, S3, is located after the karst system and after the Vauclusian spring (Appendix VII, Photo VII.1). At this point, the size of the entire catchment is 476 km².

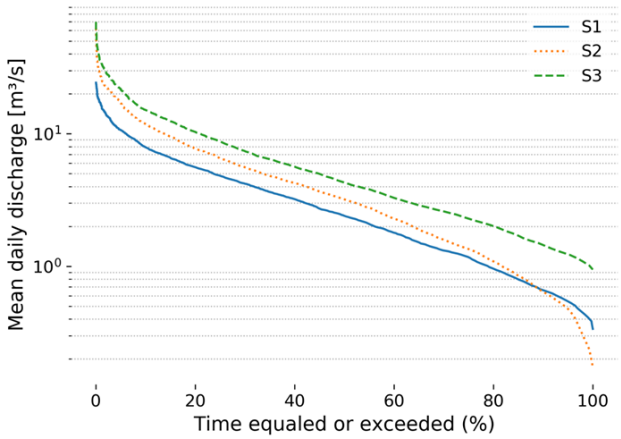

Figure 3-3: Flow duration curves for the streamflow data before (S1), inside (S2, and after (S3) the LKS (see Figure 3-1).

Each of the three time-series contains 2,557 daily records averaged from hourly data. Their respective flow duration curves (Figure 3-3) highlight the water loss during the low flow at station S2 due to percolation through the river bed. Indeed, the upstream low flow above 90% (Q90) at S1 is higher than the downstream station S2. This effect is no longer observable at S3, where the water is returned by the Vauclusian spring. The spring discharges most of S3's extreme low flow. In terms of recession, considering S2 (a perched stream), and S3 (characterized by a constant inflow-storage-discharge relationship), the prevailing hypotheses behind recession analysis are certainly not valid for these two locations. Hence, the setup of these three consecutive stations is particularly suited to study nonlinearities in streamflow recession arising from a complex hydrogeological context (e.g., Clark et al., 2009), primarily using empirical methods that make no hypotheses on the form of nonlinearities, not as Eq. 1.5 does. The three time-series are used for this purpose in Chapter 4.

3.3.2 Electrical Resistivity Tomography (ERT)

The ERT experiment allowed collecting ERT datasets daily between 2014 and 2017, which still represents, to the best of the author's knowledge, the longest, high-resolution ERT monitoring experiment conducted in a karst environment. The electrodes are permanently installed along a line of 48 electrodes at 1-meter intervals. The line starts at the bottom of a sinkhole and goes all the way to the top of a flat limestone plateau (Figure 3-2, B, Appendix

VII, Photo VII.6). Most of the electrodes are permanently buried at shallow depth, while the first six electrodes are directly attached to the outcropping limestone. Measurements were carried out first via an ALERT system (Kuras et al., 2009) and then with a Syscal Pro (Iris Instruments) and include dipole-dipole and gradients protocols. Data quality was assessed via reciprocal measurements. Resistivity models were processed using BERT (Günther et al., 2006; Rücker et al., 2006) using a time-lapse inversion scheme with a reference model (Figure 3-4.a). For a more detailed presentation of the measurements and the inversion aspect, see Watlet et al. (2018b). The dataset is also accessible from the supplementary data on a Zenodo repository (Watlet et al., 2018a).

The time-lapse ERT dataset (Figure 3-4) is obtained from dipole-dipole arrays. The spatial grid consists of 1558 cells. Each of them is assigned to a resistivity time-series defined on 465 daily time-steps defining the temporal dimension N of the dataset. From 2015-04-13, 389 measurments were obtained between 10 and 11 p.m, and therefore integrate most of the rainfall occurring on the same day. Before, the measurments were mostly obtained at noon (55 of them), and between 0 a.m. and 9 a.m. for the rest of them. Different acquisition times can generate dynamic noise that can make causal inference more difficult in Chapter 6. Also, several gaps occur throughout the dataset (Figure 3-4.d). Such gaps are inherent to field measurements and should be accounted for when searching for semi-automated tools to support the interpretation of time-lapse ERT results (Chapter 5 and 6).

This dataset has two main strengths: (i) it images a complex fractured limestone area and therefore shows a vast range of resistivity patterns both spatially and temporally, and (ii) it is a 2D ERT profile case study, which is an advantage when testing several clustering approaches. These two aspects seem ideal to explore different clustering methods in the context of identifying geological features with distinct hydrological patterns, i.e., hydrofacies. In Chapter 5, this objective is pursued by grouping together similar resistivity time-series (Figure 3-4.d) using unsupervised time-series clustering methods. Clustering will reduce the spatial dimension of the model from 1558 cells to a small number of clusters representing hydrofacies. Hence, besides hydrofacies detection, it is a task of dimensionality reduction (section 1.1.2.2).

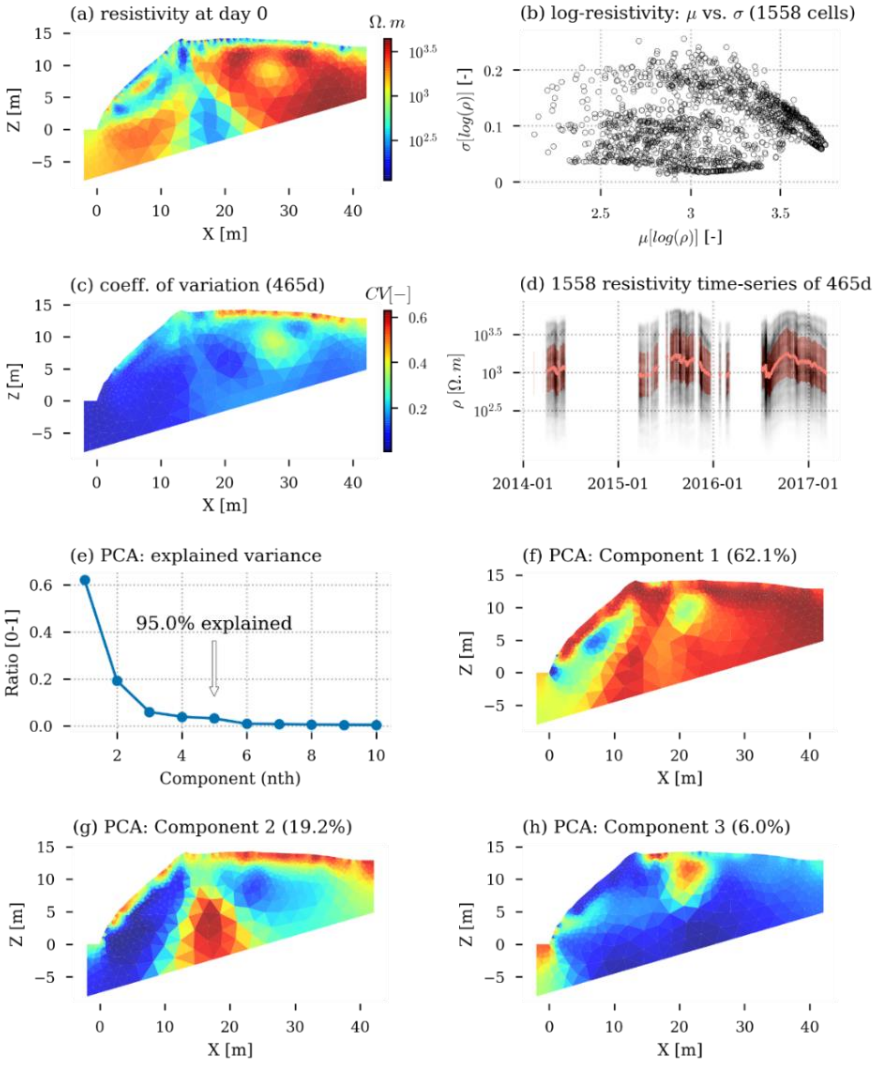

Figure 3-4: Time-lapse ERT model of the Rochefort cave subsurface. (a) Inverted resistivity ρ for the reference model at day 0. (b) Scatterplot of the mean log-resistivity $\mu[log(\rho)]$ and standard deviation of log-resistivity $\sigma[log(\rho)]$ computed over the 465 days for the 1558 spatial cells. (c) Coefficient of variation ($CV = \sigma(\rho)/\mu(\rho)$) of the resistivity computed over the 465 days. (d) Inverted daily resistivity time-series (log-scale). The red line is the mean time-series, with the shaded red areas representing the interquartile range. Missing data are left blank. (e) Variance explained by the principal component analysis (PCA) of the z-standardized resistivity dataset (see Eq. 5.1). (f to h) First, second, and third principal components associated with (e).

Furthermore, all time-series associated with a hydrofacies can be averaged into one single time-series representative of the hydrofacies resistivity dynamics. The averaging would yield a reduced set of time-series that could be used to investigate the causal relationships in the Lorette cave with causal inference methods (Chapter 6). A first visual appreciation of the spatial patterns can be appreciated by examining the representations of the dataset reporting the coefficient of variation (Figure 3-4.c) and the three first principal components (Figure 3-4, f to h). The components result from the principal component analysis of the z-standardized data (i.e., a decomposition of the correlation matrix) and explain 87.3% of its variance.

In addition, Watlet et al., 2018b performed a supervised classification of the dataset (Figure 3-5). The classification was based on the resistivity patterns (Figure 3-4) while accounting for external information: an in-situ borehole, geological observations, and a 3D model from a UAV-based photoscan (Triantafyllou et al., 2019) performed in the cave. The interpretation resulted in a segmented classification of the model into the eight zones shown in Figure 3-5. The highly resistive zones under the plateau (Figure 3-4.a) were interpreted as low porosity limestone (Figure 3-5, zones D & F). More conductive patterns were attributed to either the soil (Figure 3-5, zones A & C), the karstified limestone areas (Figure 3-5, zones B & E), or a zone of increased fracture intensity with a strong dip in the middle of the image (Figure 3-5, zone H). Lastly, zone G presents a low and relatively constant resistivity (see Figure 3-4) related to the presence of clayey limestone. The classification was limited to the upper model because the experts took into account the loss of resolution in the lower part.

Based on the PCA first component (Figure 3-4.f), the dynamic high resistive limestone zone F is correlated with the clayey limestone (Figure 3-5, zone G). The other massive limestone zone D (Figure 3-5) is correlated with the rest of the model (Figure 3-4, f to h). On the second component (Figure 3-4.g), the superficial zones A to C appear more clearly. The porous limestone area E is also identifiable in blue tones, as well as a spot of higher conductivity on the reference model (Figure 3-4.a), or using the coefficient of variation (Figure 3-4.c). The patterns of the third component (Figure 3-4.h) are mostly redundant with the first one, except for the lower left part of the model. However, that area was not considered in Figure 3-4.a since it consists of extrapolated resistivity values given that the first electrode is located at the origin of the X and Z coordinates.

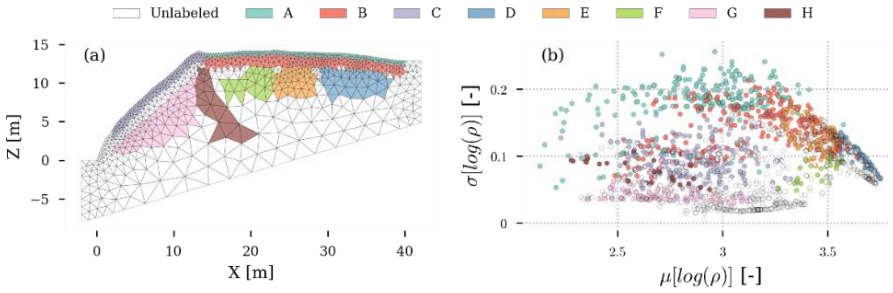

Figure 3-5: Expert-based classification of the time-lapse ERT model from Watlet et al. (2018b). (a) Spatial zonation of the groups. (b) Scatterplot of the mean log-resistivity $\mu[log(\rho)]$, and the standard deviation of log-resistivity $\sigma[log(\rho)]$ for the 1558 resistivity time-series. Colors correspond to the groups identified in (a).

3.3.3 Other Time-Series

Figure 3-6 shows additional environmental time-series considered for causal inference in Chapter 6. Potential evapotranspiration data (ET) are representative of the PAMESEB station of Jemelle (Figure 3-1). Rainfall data (RF), and drip discharge data within the cave (P1 to P3), and relative gravimetry (RG), and groundwater level (GL), are those obtained from the sensors reported in Figure 3-2 with the same codes. Atmospheric pressure data (AP) comes from a barometer associated with the gravimeter (RG).

ET consists of a daily average of potential hourly evapotranspiration estimated with the Penman-Monteith FAO-56 method (Allen et al., 1998) using PAMESEB agrometeorological data. Potential evapotranspiration does not represent actual evapotranspiration but the evapotranspiration of grass in the absence of water stress. Similarly, RF data are a daily average of hourly rainfall data. P1, P2, and P3 are daily means of the percolation rate (obtained from Watlet et al., 2018a). AP and RG are a daily average of hourly data obtained from the ROB. The RG signal is corrected for tidal, atmospheric, polar motion effects, and instrumental drift (see Watlet et al., 2020). Note that in contrast with Watlet et al. (2020), the hourly gravity measurements are linearly corrected considering an admittance (i.e., a linear slope coefficient) of -3.3 $nm.s^{-2}.hPa^{-1}$. This theoretical factor considers the effects resulting from changes in the density of the atmosphere (~ -4 $nm.s^{-2}.hPa^{-1}$) and the density of the earth's crust induced by atmospheric pressure (~ 1 $nm.s^{-2}.hPa^{-1}$). However, some linear correlations between RG and AP may be observed, given that this factor is theoretical. Hence, Watlet et al. (2020) reported and relied on an

actual admittance of -2.95 $nm.s^{-2}.hPa^{-1}$. Finally, GL is a daily average of 15-min interval data of water level provided by UNamur. The water level is inferred from water pressure and was corrected for atmospheric pressure (by UNamur, using another barometer than the one associated with AP). Summary statistics for the entire dataset are presented in Table 3-1.

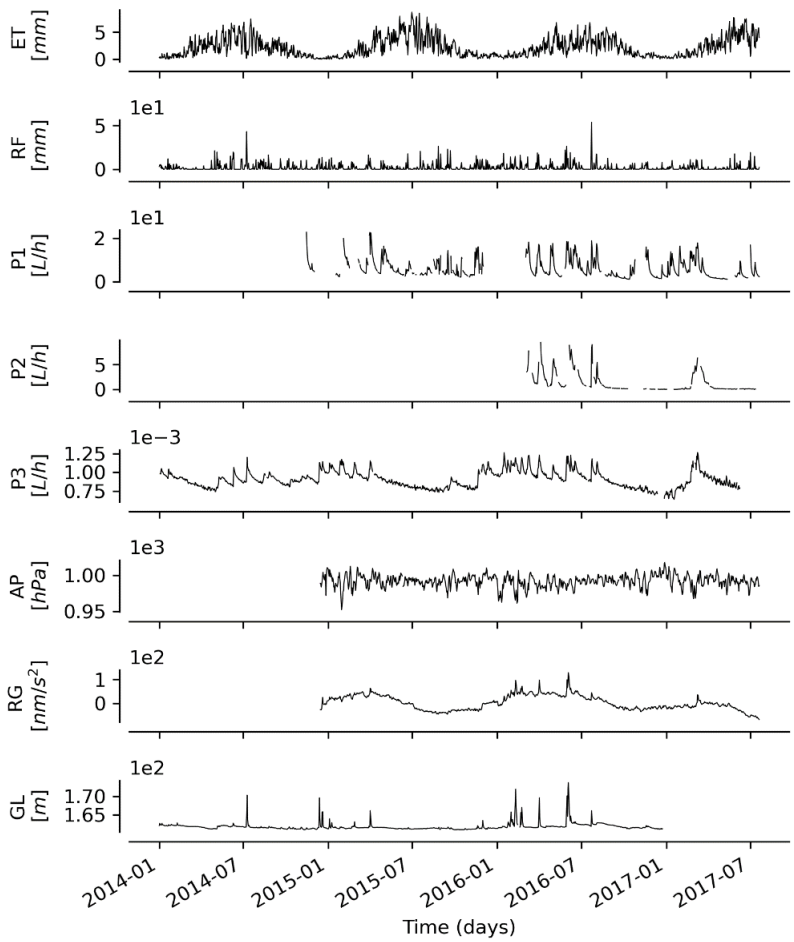

Figure 3-6: Environmental daily time-series used in Chapter 6. EP: potential evapotranspiration; RF: rainfall; P1 to P3: drip discharge data within the cave; AP: atmospheric pressure; RG: relative gravimetry; GL: groundwater level.

Table 3-1: Summary statistics of environmental time-series.

	ET [mm]	RF [mm]	P1 [L/h]	P2 [L/h]	P3 [L/h]	AP [hPa]	RG [nm/s^2]	GL [m]
count	1297	1297	718	366	1223	947	948	1084
mean	2.18	2.01	6.13	1.30	9.06e-04	991.73	3.39	161.86
std	1.81	4.21	4.33	1.92	1.11e-04	8.63	29.92	1.06
min	0.03	0	1.05	3.49e-03	6.35e-04	952.74	-66.41	161.04
5%	0.19	0	1.49	2.52e-02	7.53e-04	976.24	-38.38	161.18
10%	0.30	0	1.91	4.64e-02	7.76e-04	980.66	-31.99	161.24
25%	0.67	0	3.00	8.90e-02	8.21e-04	987.18	-21.14	161.41
50%	1.63	0.10	4.57	0.27	8.99e-04	992.27	-0.19	161.62
75%	3.30	2.08	8.46	1.73	9.75e-04	997.50	29.89	161.96
90%	4.97	6.24	12.74	4.20	1.06e-03	1001.51	42.66	162.42
95%	5.79	10.69	15.33	5.48	1.11e-03	1004.75	47.20	162.87
max	8.56	53.83	22.66	9.45	1.26e-03	1017.15	127.66	173.67

Regarding the dynamics, the P1 discharge, due to the existence of fast preferential flow (see section 3.2; Poulain et al., 2018), exhibits the highest rate and reactivity between the three discharge series. As P3 is monitoring a stalactite drip discharge, the rate is four orders of magnitude below P1. The residual surface relative gravimetry (RG) reflects temporal changes in the hydrological mass balance of the cave system. RG integrates water mass variations from the surface (~220 m.a.s.l.) according to an acquisition cone that extends down to the saturated zone (~150 m.a.s.l.). For the Rochefort cave system, this cone has a base with a radius of approximately 750 m. The angle effect, in addition to the fact that the gravimeter is installed in a shelter, can create an umbrella effect such that rain infiltration can only be detected after a certain period. GL is the groundwater level monitored in the cave of Nou Maulin. Although piezometric data are available within the Rochefort Cave (Poulain, 2017; Watlet et al., 2020), data from the cave of Nou Maulin were preferred because more correlated to the gravity signal. Although not apparent in Figure 3-6 because of the magnitude of flood peaks, GL presents a continuous dynamic, while series in the Lorette cave are intermittent as the pressure sensor is not always submerged. GL also captures well the flood peaks due to the overflowing of the Lhomme river into the karst system. The same peaks are observable and captured in the RG signal. However, the relaxation of the gravity signal is slower than the recession of the GL, which allowed Watlet et al. (2020) to conclude that the RG response to floods includes internal storage in unknown karst voids and to estimate their volume

and location. Temporal dependencies and causalities between these time-series are treated explicitly in Chapter 6, including resistivity time-series (section 3.3.2).

Chapter 4 Measuring karst-induced nonlinearity on river recession dynamics

"Never acknowledge something for true unless you evidently know it as such."

Descartes

Foreword

I could have quoted the integrality of Descartes' method, as every sentence reminds me of the laborious path behind this chapter. Indeed, Descartes'method testifies to the difficulty of following your own path of reason, a path that has not been taken until then or abandoned for a while. Initially, there was no intention to study the dynamics of the Lhomme River recession profoundly (section 1.2.1), and I went into the simple methods of recession analysis and hydrograph separation to get a general idea of the hydrological functioning of the karst river system (section 1.1.1). However, I quickly realized that these methods were possibly contraindicated, given the violation of their assumptions when applied on a perched stream, heterogeneous aquifer with permanent recharge. At the same time, I was studying the theory of nonlinear systems (e.g., Figure 1-4). As I learned more and more, I had doubts about their application in hydrology, especially with strong hypotheses of determinism and thus the absence of noise, time-series of infinite length, low dimensionality, or absence of intermittency. Gradually, I began to see that the theory of nonlinear systems applied to the case of the recession could be a stimulating and compatible case study to overcome the parallel doubts that I cultivated concerning recession and nonlinear systems theories.

This path undertaken on doubts was very difficult to clear. I started my analyses in summer 2017 and aimed to cover both the subject of recession analysis and flow separation. I presented the first poster on the subject in April 2018 (Delforge et al., 2018). Some of the patterns presented in the poster were later found to be non-robust. The remaining one (see Appendix IV) was still part of the first version of a paper submitted to Water Resource Research at

the end of 2018. As the paper had a major revision for being too cumbersome and skeptical about the recession theory, it was completely rewritten from scratch and submitted during summer 2019, dropping the section on hydrograph separation to focus on the recession analysis and various illustration of the potential application of the nonlinear theory to the field of recession analysis. The paper experienced a second major revision because, although considered attractive, still too scattered and not convincing enough. As a result, the latest revision, which was finally accepted (Delforge et al., 2020) and is adapted in this chapter, has focused on analyzing the nonlinearity of the recession, and included a robust sensitivity analysis supporting the results together with a more detailed interpretation to relate patterns to catchment characteristics (i.e., causality in the sense of Chapter 2). Elements withdrawn from the previous version are still mentioned in the perspective of the chapter.

From a philosophical point of view, the chapter questions nonlinearity as an indicator of the geomorphological complexity of a watershed (section 1.1.2.3). It compares a parametric definition of nonlinearity provided by the B&N model (section 1.1.1, Eq. 1.5) with the empirical definition of nonlinearity, i.e., the sensitivity to initial conditions, provided by a nearest-neighbor regressor (EDM-Simplex) rooted in the theory of nonlinear dynamical systems (Figure 1-4). This is thus a reflection on the form of nonlinearity (d8, Appendix III) and the adequacy of the definition with respect to the system's perceived complexity (d34, Appendix III). Indeed, since nonlinearity is an indicator of a system's organization, it is expected to be more important for a complex watershed, i.e., downstream of the karst for the LKS system. Besides, recession analysis applies to recession segments that can be interpreted as the material cause for the approach (d7, Appendix III). Through sensitivity analysis, the paper also analyses the impact of the definition of recession segments on the conclusions regarding nonlinearity. In the end, nonlinearity does indeed depend on the choice of the definition of nonlinearity (d8) and the recession segments (d7). Although nonlinearity is related to the mechanism (section 1.1.2.3), the chapter does not discuss the mechanisms in the sense of hydrological processes (d9, Appendix III) and argues that the B&N model has become empirical (d9→d8) in the context of recession analysis because often applied outside the scope of its hypotheses. The final cause (d10, Appendix III) is not much discussed in terms of operational purposes because the paper is very methodological, and the practical usefulness of the model (EDM-Simplex) is only demonstrated in terms of the quality of the forecasts.

Abstract

For more than a century, the study of streamflow recession has been dominated by seemingly physically-based parametric methods that make assumptions on the nonlinear nature of the hydrograph recession. In practice, several studies have shown that various degrees of nonlinearity occur in the same time-series and that parametric methods can underfit nonlinear recession patterns. As a result, these methods are often applied empirically and individually to each recession segment instead of modeling all recession points. The chapter proposes a parsimonious data-driven model, EDM-Simplex, with two objectives: forecasting recession and characterizing its nonlinear behavior. The new model is evaluated through global sensitivity analysis (GSA) applied to three distinctive hydrograph series from a heterogeneous karstic catchment (section 3.3.1).

The results show excellent 1-day-ahead forecasting performance (median Nash and Sutcliffe efficiency > 0.99, Eq. 1.4) for all time-series with four recession extraction methods. The sensitivity analysis also showed that empirical nonlinearity, that is, sensitivity to initial conditions, is best estimated through the absolute forecast performance and its decline over time. This indicator leads to different interpretations of nonlinearity compared to previous methods but is just as sensitive to the choice of recession extraction method. In particular, when forecasts were made for recession segments containing early stages of recession or flow anomalies, the upstream recession was significantly more linear than the downstream recession hydrographs affected by the karst. Consequently, the results support future research to interpret observed nonlinearities as a function of the catchment hydrological states for better integration of empirical, physical-based, and operational approaches to recession analysis.

4.1 Introduction

River flow integrates the systemic dynamics of the entire watershed. By definition, the hydrograph of a river shows the streamflow variation in time, shifting between flood and recession periods, depending on whether the watershed is under the influence of a rainfall event or not (Figure 4-1). Recession analysis methods focus on the decreasing river baseflow in the absence of rainfall. Applications of recession analysis include low flow forecasting, evaluation of aquifer recharge or other hydrological mass balance components, estimation of the aquifer hydraulic properties, water supply planning for human needs, evaluation of ecosystems sustainability, the study of climate or anthropogenic impact on water resources, water quality monitoring, and waste management (Dewandel et al., 2003; Smakhtin, 2001; Tallaksen, 1995; Wittenberg and Sivapalan, 1999).

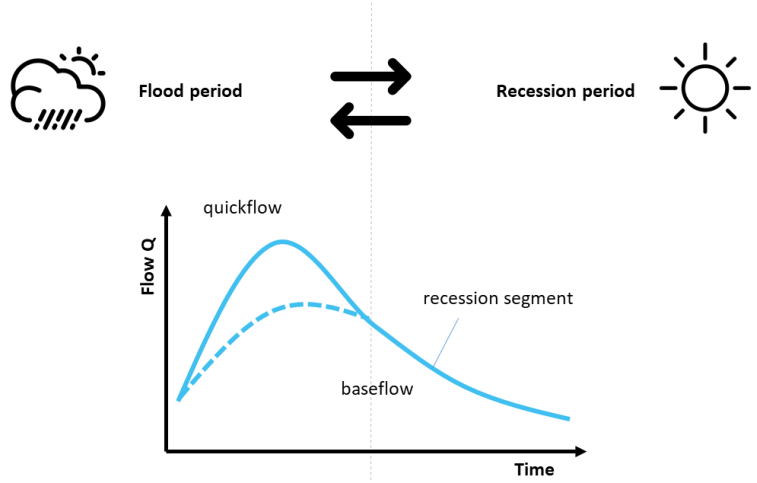

Figure 4-1: Schematic representation of the hydrograph. Source: icons are from the noun project (thenounproject.com).

Baseflow recession is seen as the simplest dynamic process by assuming that antecedent rainfall has no longer an influence on the decreasing dynamic of streamflow and only depends on its own state. The Brutsaert and Nieber (1977) (B&N) model of streamflow recession incorporates this assumption explicitly by relying on streamflow as the only variable, with only two parameters ($-dQ/dt = aQ^b$, Eq. 1.5). For more than 40 years, the B&N model has been considered as a reference model for the study of the daily

streamflow recession. Eq. 1.5 is physically based and related to groundwater hydraulic theory, in particular to the storage-discharge relationship developed by Boussinesq (1903), among others (see the reviews of Dewandel et al., 2003; Troch et al., 2013).

Boussinesq provided an exact solution of the diffusion equation for an unconfined homogenous aquifer sitting on a horizontal impermeable layer, which allows inferring hydraulic properties of the aquifer. Boussinesq's solution suggests that the storage-discharge relationship for this idealized aquifer follows a nonlinear quadratic form, which corresponds to a value of b = 1.5 (Eq. 1.5). Although some experimental studies report b values close to 1.5 (e.g., Troch et al., 1993; Wittenberg, 1994), others provide observed b values ranging between 1 (i.e., linear recession) and 3 (e.g., Chapman, 1999; Wittenberg, 1999). This suggests that field observations are often incompatible with the assumptions of Boussinesq groundwater hydraulic theory.

There are many reasons for such variations on the range of the nonlinear parameter b. For example, the early recession is expected to exhibit a higher b exponent (Brutsaert and Nieber, 1977). Another possible answer considers the aquifer heterogeneities affecting the saturated hydraulic conductivity. Some papers report on these effects, assuming either horizontal or vertical anisotropy (Harman and Sivapalan, 2009b; Rupp and Selker, 2005, 2006), which result in an increase of the nonlinear behavior. Another reason is that streamflow recession is related to spatial heterogeneity of the catchments (Troch et al., 2013). Under specific conditions, the subsurface discharge from hillslopes follows a linear relationship (Berne et al., 2005; Harman and Sivapalan, 2009a). Conceivably, the parallel and additive contribution of different linear hillslope subsurface discharges to streamflow could be related to the nonlinear exponents of the recession (Clark et al., 2009; Harman et al., 2009). Biswal and Marani (2010) relate the nonlinear properties of baseflow regimes to the morphology of the contributive channel network. Other authors observed a seasonal variation of b by analyzing individual recession events with the B&N model. These variations were attributed to change in the catchment evapotranspiration dynamics (Shaw and Riha, 2012), or, from a systemic point of view, to the combined effects between evapotranspiration, antecedent soil moisture, historical recharge conditions, and catchment geomorphological features. It was suggested that low-flow contributive areas vary over time (Karlsen et al., 2019; McMillan et al., 2011; Shaw et al., 2013).

Bogaart et al. (2016), based on a comparative analysis of 200 catchments in Sweden, found long-term trends showing an increase in recession nonlinearities. They related them mainly to changes in the land cover.

Presently, the processes and explanatory variables to be included in an improved model are not well understood and are not likely to be incorporated rapidly in a new generic model as they appear to be generally related to a systemic and specific behavior of catchments. Some of the aforementioned studies clearly show that the B&N model missed patterns in the nonlinear behavior of low-flow regimes, and this has motivated the sequential, piecemeal application of the B&N model to individual segments (Karlsen et al., 2019; McMillan et al., 2011; Shaw et al., 2013; Shaw and Riha, 2012). For example, Jachens et al. (2020) encourage the application of the B&N framework systematically on individual recession segments for the sake of consistency. Importantly, all these studies clearly show the limits of the B&N parametric framework of low dimensionality for recession analysis. They suggest considering low flow as a higher dimensional problem that could be analyzed using pattern recognition or models of higher degrees of freedom. Moreover, different studies showed that the experimentally optimized B&N model parameters are highly sensitive to the recession extraction method, the observational error, the choice of the fitting method, and the possibility that b exponents obtained from fitting all recession points (Eq. 1.6) might be inconsistent with individual events (Dralle et al., 2017; Roques et al., 2017; Stoelzle et al., 2013). In a broader sense, although recognized as the simplest dynamic in the hydrograph, the study of streamflow recession is currently facing the same challenges as encountered in catchment hydrology at large, that is, bridging physical processes, patterns, and functional traits at all scales and beyond complexity (McDonnell et al., 2007; Sivapalan, 2006). Consequently, it is considered that the use of the B&N model on individual recession segments (i.e., by considering the two parameters of Eq. 1.5 times the number of recession segments) reduces it to an empirical model that does not respect the principle of parsimony (section 2.4.1).

The objective of this work is to propose a parsimonious and robust empirical framework to forecast the nonlinear recession and assess its nonlinearity. The method relies on the empirical dynamic modeling (EDM) framework (Sugihara, 1994; Sugihara et al., 2012; Sugihara and May, 1990; Ye and Sugihara, 2016). EDM is based on the theory of nonlinear dynamic systems, commonly referred to as chaos theory (Lorenz, 1963; Takens, 1981). The

proposed EDM-Simplex method (Sugihara and May, 1990) is a nearest-neighbor regressor for univariate time-series analysis and forecasting. The EDM-Simplex method is based on the general definition of nonlinearity, meaning sensitive dependence on initial conditions. From both operational and scientific perspectives, the chapter studies the performance of EDM-Simplex's empirical nonlinear predictions, as well as the method robustness to its variables and assumptions. For these purposes, global sensitivity analysis (GSA) is applied (Cukier et al., 1978; Saltelli et al., 2000, 2007) involving various choices of recession extraction methods and EDM-Simplex configurations. The method is applied on three streamflow data sets from three successive stations gauging the same river with an expected increase in the nonlinear behavior due to the downstream presence of a karstic system (section 3.3.1, and Figure 3-3). The consistency between the parametric B&N approach and the empirical EDM-Simplex method is further discussed together with the emerging perspective of connecting the dots between physical and empirical approaches in hydrograph recession analysis.

4.2 Theory and Methods

4.2.1 Overview and Philosophy

Since it is argued that the B&N approach has become empirical, an alternative and more parsimonious empirical approach to the usual B&N analytical framework is proposed to assess the nonlinearity of the recession: EDM-Simplex (Sugihara and May, 1990; Table 4-1). Both approaches are deterministic, but B&N is grounded in the mechanistic and parametric philosophy, while EDM-Simplex relies on an empirical formulation of determinism: recession dynamics will behave as usual under similar conditions. Accordingly, with EDM-Simplex, the future recession of recession segments is forecast by averaging the future recession of similar segments, which are supposed to reflect similar environmental conditions, in virtue of the theory of deterministic and nonlinear dynamical systems (Figure 1-4, and section 4.2.2.1 for details). EDM-Simplex belongs, therefore, to the family of nearest-neighbor regressors (section 4.2.2.2 for the detailed algorithm).

Table 4-1: Comparison between the B&N and the EDM-Simplex approach.

	B&N	**EDM-Simplex**
Philosophy	Mechanistic determinism *Recession dynamics will follow the groundwater hydraulic theory.*	Empirical determinism *Recession dynamics will behave as usual under similar conditions.*
Model	The B&N approach $-dQ/dt = aQ^b$ Or in the log-log form: $\log\left(-\frac{dQ}{dt}\right) = \log(a) + b\log(Q)$	EDM-Simplex nearest-neighbor regressor forecasting the future of recession segments based on the future of similar recession segments.
Nonlinearity	The nonlinear exponent b	Sensitive dependence to initial conditions: - performance decay for a predictive horizon tp; - the length of recession segment m (or embedding dimension); - number of nearest neighbor segments k;
Other parameters	The recession constant a	See Table 4-3
Recession extraction methods	Section 4.2.3 BRU, EDM, VOG, KIR	Section 4.2.3 BRU, EDM VOG, KIR (+ a truncation parameter h, Table 4-3 and Table 4-4)

The theory of nonlinear dynamical systems has been extensively applied to streamflow records (reviewed in Sivakumar, 2017). Besides, several studies demonstrate the applicability and good performance of regressors based on approximation from the local neighborhood on the forecasting of daily streamflow hydrographs (e.g., Islam and Sivakumar, 2002; Kember et al., 1993; Khatibi et al., 2012). However, none have been specifically applied to predict and study recession dynamics alone, especially under the light of the theory of nonlinear dynamical system. Yet, applying the chaos theory to entire streamflow series to infer chaotic properties is challenging due to the high dimensionality of hydrograph dynamics, the presence of intermittent processes, the shortness of time-series, and the presence of noise (see Koutsoyiannis, 2006b; Sivakumar, 2017). Indeed, applied to the prediction of floods, a nearest-neighbor regressor forecasting the future of flood events from the future of similar flood events is likely to perform poorly because the dynamic is forced by an exogeneous, random, or high dimensional variable, i.e., rainfall.

Regarding recession only, it is assumed that recession, while not chaotic in the sense of apparent randomness, is a nonlinear, recurrent, and low dimensional process, less noisy and uncertain than the flood hydrograph, and therefore congruent with the theory of nonlinear dynamic systems. Recession is no longer considered as the sustained decreasing streamflow but as the low-dimensional, deterministic, and recurrent part of the hydrograph.

The modeling philosophy is the following. One can imagine the dynamic of a river as the one of a reservoir irregularly disturbed by a stochastic component such as rainfall but whose intrinsic dynamics tend toward more determinism as recession progresses and the influence of rainfall diminishes. This image is not so far from that of a ball revolving in an irregular bowl (watershed) naturally attracted by the lowest point, where an unpredictable hand (rainfall) would prevent it from reaching its focus (Figure 4-2). In either view, system disturbances are expected to be relatively unpredictable, dynamically complex, and followed by a relatively simpler motion when a purely deterministic dynamic takes over, a dynamic guided by the physical and stable structure or organization of the system. Focusing on recession allows inferring about "the bowl," that is, catchment geomorphology and organization, with fewer concerns about the rainfall-runoff response that is either random or of much higher dimensionality and for which EDM-Simplex may not be a right tool or would demand much longer time-series (see also Figure 1-5, section 1.1.2.3 and 1.1.2.4).

Accordingly, EDM-Simplex assumes stationarity, not in the statistical sense, but rather a structural one that assumes that the unknown set of dynamical equations remains unchanged, including its parameters (Kantz and Schreiber, 2003). Pictorially, the shape of the bowl has to remain constant (Figure 4-2).

Pragmatically, there are methodological considerations to be discussed. In general, the forecasting performance of the EDM-Simplex model is used as a basis for inferring geomorphological complexity. The performance metric is the Nash and Sutcliffe (1970) efficiency (NSE, Eq. 1.4).

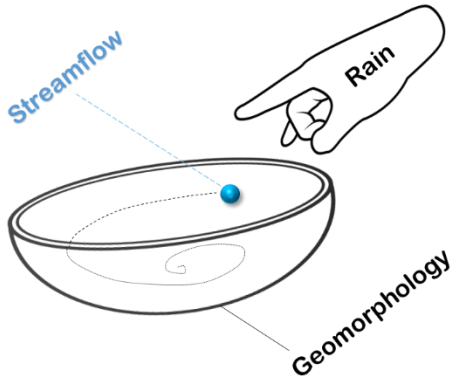

Figure 4-2: Modelling philosophy behind an empirical, deterministic, and nonlinear philosophy for streamflow recession analysis. Recession is seen as the deterministic part of the streamflow hydrograph. While streamflow is under the influence of rainfall (the hand), its dynamic trajectory is unpredictable because rainfall is random. As the influence of rain diminishes, i.e., during recession periods, the trajectory taken by the hydrograph becomes more and more deterministic and reflects the characteristics of the geomorphology (the bowl), whatsoever its shape, i.e., with no physical assumptions. Accordingly, a nearest-neighbor empirical model such as EDM-Simplex can be used to assess the nonlinearity of recession induced by the geomorphology. Furthermore, EDM-Simplex can extract deterministic points in the hydrograph where trajectories are consistent and provide good forecasts, that are, expectedly, recession points. Source: icons are from the noun project (thenounproject.com).

Hypothetically, three EDM-Simplex parameters relate to the observed dynamics' nonlinearity or complexity (Table 4-1): the time to prediction tp, the length of recession segments in days m (or embedding dimension), and the number of necessary similar recession segments k. As highlighted by Lorenz (1963), nonlinear dynamics exhibit a higher prediction decay through time. Hence, reporting the EDM-Simplex NSE performance decay against the prediction horizon tp is a way to measure nonlinearity. According to the theory (section 4.2.2.1), the dimension of a system, i.e., its complexity (section 1.1.2.2), can be approached by the parameter m, which in our case translates the length of the recession segments. Intuitively, the more complex the geomorphology of the watershed, the more days of recession (e.g., 1, 2, or 3 days) should be considered for optimal NSE performance based on similar segments. Finally, the number of nearest neighbors k is potentially a gateway to assess nonlinearity empirically. It is assumed that the more nonlinear the dynamic is, the better the prediction will be by considering the close neighborhood, that is, a reduced number of nearest neighbors. A reduced set of neighbors is supposed to capture better environmental conditions'

specificity, which matters if the dynamic is nonlinear. Still, Sugihara and May (1990) suggest using at least $k = m + 1$ to close the dynamic according to its dimension. These $m + 1$ neighbors are called a simplex, which gives the name to the original algorithm. EDM-Simplex is an ensemble model relying on bootstrapping. Accordingly, the size of the bootstrap samples L is supposed to have an influence on k, as it will affect the density of nearest-neighbor recession segments in the analysis.

Thus, potentially, tp, m, and k are related to the nonlinearity of a process. For an actual case such as recession, this would be related if these parameters are sensitive while investigating different recession hydrographs associated with catchment showing different hydrogeomorphological properties. If they are sensitive, they are allowed to be used as indicators of nonlinear or complex behavior. Given the number of EDM-Simplex, they might be a preferred set-up of parameter to apply EDM-Simplex. Although less related to nonlinearity, other parameters and factors involved in the EDM-Simplex algorithm (such as L and others; section 4.2.2.2) may also affect the outcome and turns to be important. Furthermore, the results may be influenced by the recession extraction methods that spot the points that will be forecast by the algorithm, as it is the case for the B&N framework (e.g., Stoelzle et al., 2013). Hence, various recession extraction methods are considered in this issue (section 4.2.3), including one that involves EDM-Simplex and its philosophy (Figure 4-2, see also Appendix V). In total, the parameters of the EDM-Simplex algorithm and the recession extraction methods are all subject to a robust global sensitivity analysis (section 4.2.4) to demonstrate the use of EDM-Simplex for the analysis of recession nonlinearity, its predictive capabilities, and to define recommendations for future applications. The predictions are applied to the three stations in the Lhomme streamflow dataset (section 3.3.1). The Lhomme catchment is a very heterogeneous watershed offering the opportunity to relate the complexity of the watershed to non-linear recession patterns (section 3.1).

4.2.2 EDM-Simplex Model

4.2.2.1 Takens's Embedding Theorem

The theory of nonlinear dynamical systems applied to univariate time-series analysis (Kantz and Schreiber, 2003) commonly relies on Takens's (1981) embedding theorem, which suggests that the dynamic of a system represented

by its state space M can be reconstructed into a pseudo state space M_X based on one single time-series X_t and its lagged coordinates:

$$M_X(\tau, m) = \{X_t, X_{t-\tau}, \dots, X_{t-(m-1)\tau}\} \qquad Eq.4.1$$

where m is the embedding dimension and τ the embedding delay. M is the D-dimensional space reporting the states of a system with respect to the D explanatory variables driving its trajectory. Nonlinear deterministic systems draw unique and recurrent trajectories, meaning that the dynamic revisits closely the same portion of the state space. Accordingly, the trajectories delineate a confined state space, that is, a basin of attraction whose geometry is called an attractor (see Figure 1-4). The reconstructed attractor M_X from a single variable endogenous to M is a proxy to investigate the dynamical complexity, the dimension, and the nonlinear properties of the original attractor M. Takens's theorem further relates the system dimension D to the reconstructed dimension m by the following inequality: $m > 2D$ will always yield to a faithful reconstruction in the absence of noise. Still, m can be smaller than $2D$. It depends on the geometrical complexity of the reconstructed trajectories.

The theorem assumptions hold for low-dimensional deterministic systems. The deterministic hypothesis ensures a bidirectional and one-to-one map between the states of a system and a single variable trajectory. As for an ordinary differential equation, since the trajectory of a single variable depends on the other variables (system state), the delayed embedding aims at retrieving the system states from the trajectory of one of these variables. In practice, a reconstructed attractor gathers the pseudo-states, which are not physical states of the system, but qualitatively allow to estimate when the system is in similar states and to capture the patterns of its trajectories. It is these principles that allow us to make the hypothesis that two segments or trajectories of similar recession reflect two similar hydrological states of a watershed.

In hydrology, the idea behind embedding is also similar to the concept of "doing hydrology backward" (Kirchner, 2009), which consists of reconstructing the mass balance components from streamflow records regarding catchment as simple dynamical systems. In particular, the recession plot (Brutsaert and Nieber, 1977) reporting streamflow recession against its derivative (Eq. 1.5) is an alternative embedding strategy (Packard et al., 1980). The B&N model (Eq. 1.5), however, assumes that recession dynamics can be

captured in two dimensions. Hence, investigating m through Simplex forecasting hypothetically offers the opportunity to investigate this assumption.

4.2.2.2 Forecasting Algorithm

Relying on state space reconstruction (Eq. 4.1), the EDM-Simplex algorithm (Sugihara and May, 1990), as all forecasting methods in the EDM framework (Sugihara, 1994; Sugihara et al., 2012; Ye and Sugihara, 2016), is a nearest-neighbor regressor performing data-based forecasts. EDM-Simplex is specific to univariate time-series forecasting and is based on local approximations involving a minimal set of nearest neighbors. Figure 4-3 shows the flowchart of the EDM-Simplex algorithm.

For a time-series X_t, EDM-Simplex outputs a matrix of forecast $\hat{X}_{REF+tp}$, where REF are the user-defined time indices of reference and tp the time to prediction. In this study case, X_t represents a daily streamflow series of length N. The reference indices REF are the hydrograph points tp time-steps before the forecasted recession points. The recession points are identified using a recession extraction method (see section 4.2.3). X_t is first embedded in M_X using Takens's theorem (Eq. 4.1). Since the recession analysis considers continuous segments of daily recession, X_t represents a daily streamflow series of length N, τ is assumed to be 1, and M_X is, therefore, a trajectory matrix of dimension $(N - m + 1 \times m)$. EDM-Simplex iterates through the N_{REF} single reconstructed states denoted $\dot{x}_i$ to represent a single row $\{x_i, x_{i-1}, \dots, x_{i-(m-1)}\}$ or state of $M_{X,REF}$. Every single state $\dot{x}_i$ yields to N_{SAM} forecasts $\hat{x}_{i+tp}$ using nearest-neighbor distance-weighted regression on N_{SAM} bootstrapped samples $M_{X,L}$ of size L. Therefore, $\hat{X}_{REF+tp}$ has a dimension of $(N_{REF} \times N_{SAM})$. Each set of L distinct states in $M_{X,L}$ is sampled from $M_{X,LIB}$, a subset of M_X constrained by the user-defined time indices LIB, in which EDM-Simplex ensures to exclude a window defined by the tw input around the index i of the state of reference. The time exclusion window tw is the Theiler (1986) window, which allows excluding from the library points that would be neighboring due to their temporal proximity. This control mechanism prevents EDM-Simplex from behaving as a moving-average filter that finds its predictive potential in the autocorrelated structure of the time-series and, therefore, could confuse a recurrent deterministic process with a

stochastic process. In the case of recession modeling, tw can force recession points to be forecasted from other recession limbs, other seasons or years.

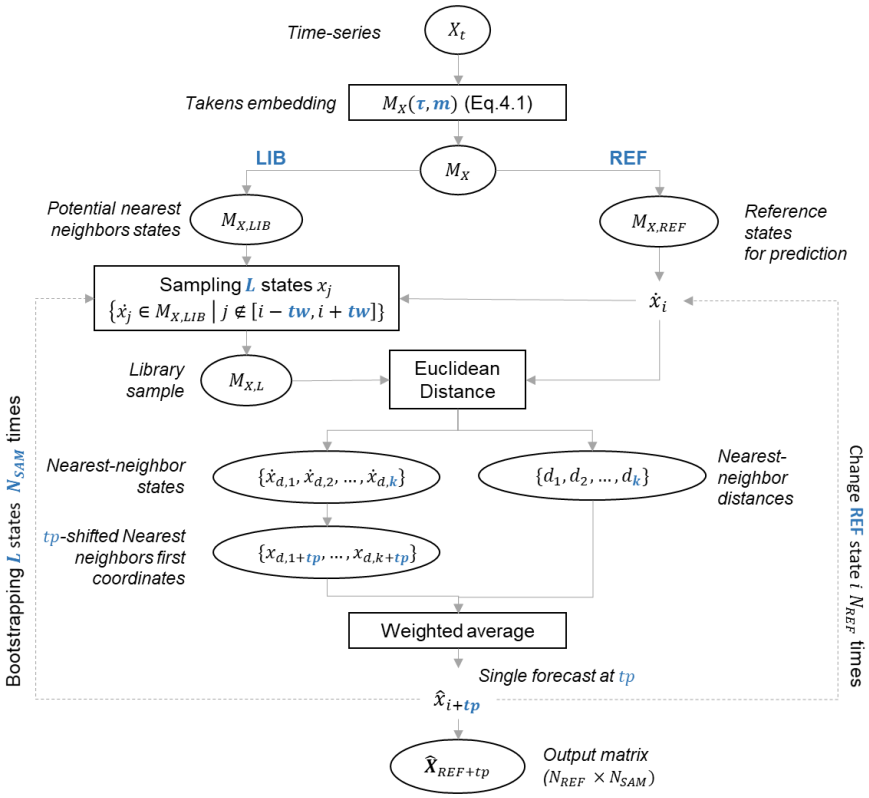

Figure 4-3: EDM-Simplex algorithm flowchart. User-defined inputs tested in the global sensitivity analysis are shown in blue and defined in Table 4-3.

To perform a single forecast $\hat{x}_{i+tp}$, EDM-Simplex computes the Euclidean distance between $\dot{x}_i$ and the library sample $M_{X,L}$ in order to retrieve a set of k nearest-neighbor states denoted $\{\dot{x}_{d,1}, \dot{x}_{d,2}, \dots, \dot{x}_{d,k}\}$ and their respective distance $\{d_1, d_2, \dots, d_k\}$. Then, Simplex identifies the states in M_X that are tp time step ahead of the neighbors. Their first undelayed coordinates $\{x_{d,1+tp}, \dots, x_{d,k+tp}\}$ are averaged to produce the forecast $\hat{x}_{i+tp}$. The averaging uses exponential weight w_j based on the distances:

$$w_j = \exp\left(-\frac{d_j}{min\{d_1, \dots, d_k\}}\right) \qquad Eq.4.2$$

Each weight is then divided by the sum of all weights so that they sum up to 1. Altogether, the user-defined inputs are summarized in Table 4-3 and section 4.2.4, as they are part of the sensitivity analysis.

The EDM-Simplex model outputs a matrix of N_{SAM} forecasts of the N_{REF} observed recession values and the median forecast. Hence, Simplex is an ensemble forecasting method.

4.2.3 Recession Extraction Methods

In this study, four recession extraction methods were used to label the recession points forecasted by EDM-Simplex. The recession segments were extracted using a series of heuristic rules reported in Table 4-2. The BRU (Brutsaert, 2008), the KIR (Kirchner, 2009), and the VOG (Vogel and Kroll, 1992) were implemented based on Stoelzle et al. (2013) and are conventional recession extraction methods. The EDM method is a new method proposed to extract recession points via EDM-Simplex based on the philosophy that recession points are the deterministic points of the hydrograph that are easily forecasted (Figure 4-2). All recession extraction methods were applied to label the raw daily data points, as they are intended to be forecasted as such using EDM-Simplex. Our implementation differs from that by Stoelzle et al. (2013) who applied a preprocessing of streamflow values by considering $Q_t = (Q_t + Q_{t-1})/2$ to keep the timing right against $dQ_t/dt = Q_t - Q_{t-1}$ while using the B&N framework (Eq. 1.5).

The BRU procedure discards the early and late parts of the recession segments and imposes that the recession rate should be decreasing through time. The KIR method is the simplest one and considers extracting all negative dQ_t/dt as $Q_t - Q_{t-1}$. Our implementation was slightly modified by considering an exclusion criterion for the first 3 days of the decreasing limb for consistency regarding the other methods and the forecasting purposes. The reasons are further motivated in section 4.2.4. The VOG method first applies a 3-day moving-average filter. Then, minimal segments of 10 days are selected. Other rules involve the exclusion of the first 30% recession points and a maximum recession rate set to 30%.

Table 4-2: Main criteria to extract recessions from streamflow data (adapted from Stoelzle et al., 2013).

Recession extraction method	Criterion	Minimum recession length [days]	Excluding exterior parts of recession segment	Excluding recession segments depending on anomalous streamflow trajectories
BRU	$dQ_t/dt < 0$	6-7	first 3-4 days, last 2 days	$dQ_{t+1}/dt > dQ_t/dt$
EDM	EDM-Simplex residuals ε $\log(\varepsilon^2/Q_t) < T$ *	2	First 3 days	Auto
KIR	$dQ_t/dt < 0$	-	First 3 days	-
VOG	Decreasing 3-day moving average	10	First 30 %	> 30 %

* Threshold value T

The EDM recession extraction method relies on EDM-Simplex (section 4.2.2) forecasting performance to identify recession points and automatically discard anomalous recession behavior (see Appendix V for an illustrated example). The EDM-Simplex model must be capable of targeting recession points that are dominated by underground baseflow discharge because it is considered as the simplest dynamics in the hydrograph. This viewpoint is consistent with the modeling paradigm adopted in this study that considers recession as the deterministic part of the hydrograph (section 4.2.1, Figure 4-2). Aksoy and Wittenberg (2011) used a similar idea to identify recession points, but based on the B&N model (Eq. 1.5) and relying on the Boussinesq's assumption that the nonlinear exponent b is equal to 1.5.

Here, the criterion is related to an error function $\log(\varepsilon^2/Q_t)$, where ε^2 are the squared residuals calculated using the median predicted values. Balancing the error with the observed discharge Q_t in the denominator allows extracting recession points for higher flows. In this application, the EDM-Simplex model was used to predict all hydrograph points and compute the error. The embedding dimension $m = 2$ was selected because of its analogy with the dimension of the Q_t versus dQ_t/dt recession plot. A prediction horizon tp of 0 days was selected since the purpose is the extraction of recession points and not the prediction of future values. The library of potential nearest-neighbor states L was indexed on decreasing streamflow values such that EDM-

Simplex was forced to perform poorly on increasing streamflow values. The sample size L was selected as 250 with a number of bootstrap samples N_{SAM} equal to 500 samples and the Theiler window tw of 10 days. The error threshold value T was set according to the 20th percentile of the distribution of $\log(\varepsilon^2/Q_t)$, respectively −10.7, −10.1 and −10.3 for S1, S2, and S3. For the given thresholds, both the mean and median of ε were found close to 0, indicating an unbiased model. The resulting points were postprocessed to fulfill the criteria of Table 4-2. First, the minimum recession length of 2 days was set to filter out the number of points that would be considered as recession points by chance, due to their excellent but fortuitous goodness of fit. Such a low value also allows having results that contrast with the BRU and VOG methods, which consider at the beginning very long recession segments. Second, all recession points that were not in the KIR set of recession points were dismissed, such that all points in the EDM method were preceded by at least 3 days of decreasing streamflow, again, for reasons that are further developed in section 4.2.4.

4.2.4 Global Sensitivity Analysis (GSA) and Sampling Distributions

A high-dimension variance decomposition-based GSA (Cukier et al., 1978; Saltelli et al., 2000, 2007) was performed to identify the most sensitive EDM-Simplex inputs that provide the best set-up to forecast recession points. The GSA also identifies the most relevant factors to interpret the complexity and the sensitive dependence to initial conditions of recession dynamics (section 4.2.1).

The study relies on the sensitivity analysis method of Sobol' (1990), which samples the factors and decomposes the variance of the outputs into first-order indices (S_i) and total-order indices (S_{Ti}). The first-order indices depict each factor's marginal contribution, while the total-order indices refer to the factor's total effect on the variance, including higher-order interactions with the other factors (Saltelli et al., 2007). Hence, the Sobol' method is particularly suited to highlight nonlinear interactions between the input factors of a model. The magnitude of these interactions can be assessed by subtracting the first-order from total-order indices (Cukier et al., 1978).

In our case, the modeling output is given by the goodness of fit of the forecast prediction quantified by the Nash and Sutcliffe (1970) efficiency (NSE, Eq.

1.4) objective function. The decomposed variance is then the variance of the NSE forecasting performances given by all EDM-Simplex input factor combinations, the Sobol' input samples, repeated for all three stations S1, S2, and S3 (Figure 3-1 and Figure 3-3). Gupta and Razavi (2018) have recently pointed out that GSAs applied to performance measures may not reflect the true sensitivity of a parametric model, which is per se independent of the observed outputs, but rather identify the most important input factors for calibration. Still, in this context, we rely on a data-driven model. The output of interest is the forecast performance, identifiability, and changes in sensitivity between application cases (Sugihara and May, 1990).

Table 4-3 (and Figure 4-3) shows the EDM-Simplex inputs selected for the GSA sampling strategy.

Table 4-3: Sampling distributions of the EDM-Simplex parameters for global sensitivity analysis.

Symbol	Description	Distrib.	Values/Bounds
τ	Simplex Embedding delay (Eq. 4.1)	Fixed	1 day
m	Simplex Embedding dimension (Eq.4.1)	Discrete	[1, 2, 3]
tp	Simplex prediction horizon	Discrete	[1, 2, 3] days
k_n	Set the number of nearest-neighbor states k such that $k = m + k_n$	Discrete	[1, 2, 3, 4]
L	Bootstrap sample size	Discrete	[10, 50, 100, 150, 200, 250, 350, 450]
tw	Theiler window of exclusion	Discrete	[3, 10, 50, 90]
N_{SAM}	Number of bootstrap samples	Discrete	[100,101, …, 1000]
LIB	Indices j defining the library of potential nearest-neighbor states $\dot{x}_j$	Fixed	$j = \{t \mid dQ_t/dt < 0, \dots, dQ_{t+tp}/dt < 0\}$
REF	Indices of states or reference $\dot{x}_i$ such that Simplex forecasts are made at $REF + tp$	Discrete	[BRU-tp, EDM-tp, KIR-tp, VOG-tp] with truncation of the segment head by h
h	Defines the size of the truncation of the head of recession segments	Discrete	[0,1,2,3,4]

For non-constant (not fixed) factors, all sampling distributions were discrete uniform. The embedding delay was considered constant and equal to 1 day since this scale is typical in daily recession analysis. Accordingly, recession points were forecasted using nearest-neighboring states $\dot{x}_j$ representing continuous hydrograph segments of length given by the embedding dimension m. The latter varied between 1 and 3 days, which was considered sufficiently large given the expected low dimensionality of the recession dynamics. Besides, EDM-Simplex is more flexible than the usual linear regression performed with the B&N framework (Eq. 1.5) since it performs local approximations, allowing modeling nonlinear patterns in the given dimension m.

The prediction horizon tp was kept below 4 days so that the algorithm remains limited to the prediction of the recession from reference states that are part of the same decreasing limb (see criteria in Table 4-2). The additional number of nearest-neighbor states k_n was set between 1 and 4. The actual number of nearest-neighbor states $k = m + k_n$ varied between 2 and 7. Given the 7-year span of the data set, the last value assumes that similar and recurrent recession points have a return period of 1 year. So effectively, $k_n = [1,4]$ explores values between the default value of 1 (Sugihara and May, 1990) and the maximum return period allowed by the testing data set. The bootstrap sample size L ranged between 10 and 450 with an increasing increment. This last value of L remains lower than the minimum size of the library of potential nearest-neighbor states LIB considered in the sensitivity analysis, which is 798 points without considering the exclusion window of Theiler tw. Preliminary testing suggested convergence in less than $L = 200$ samples generally, so the range is deemed sufficient to represent the variation of this factor. The minimal value of 3 days for tw ensured that the point being forecasted will not be selected in the library L considering the maximal prediction horizon tp of 3 days, neither points included in the reference states $\dot{x}_i = \{x_i, x_{i-1}, x_{i-2}\}$ considering an embedding dimension m of 3.

With a larger tw of 10 days, forecasts will consider nearest-neighbor states from distinct recession segments. The value of 50 days excluded 101 days such that neighboring states were picked out of the seasonal scale. Similarly, tw = 90 days excluded a half year. Given the seasonal variations in average flows, the closest neighbors are likely to be selected from different years. The number of bootstrap samples N_{SAM} was obtained from a continuous uniform

distribution and further rounded to an integer. The selected range is broad, between 100 and 1000, since there is no a priori on the value of N_{SAM} to achieve robust forecasts.

To perform EDM-Simplex forecasts on recession segments only, a careful indexing strategy is required while defining the three last inputs (LIB, REF, and h). For LIB, it is necessary that the future of the potential closest neighboring states $\{\dot{x}_{d,1}, \dot{x}_{d,2}, \dots, \dot{x}_{d,k}\}$ undergo the recession process up to the forecast horizon tp. Otherwise, EDM-Simplex will eventually output forecasts that are not strictly decreasing. Therefore, LIB identifies the potential nearest-neighbor states $\dot{x}_j$ through the time indices j as the set of time indices t satisfying the first condition $\{t \mid dQ_t/dt < 0, \dots, dQ_{t+tp}/dt < 0\}$. Thus, this condition was used to define the LIB. Stronger conditions were not deemed necessary since the closest neighbor states are automatically identified using Euclidean distance.

Regarding REF, the indexing strategy was more complicated. Indeed, forecasting recession points from recession points is particularly challenging for high tp and m due to the shortness of recession segments. For that reason, an additional input h, which truncates the head of recession segments, i.e., early recession states, was considered in the sensitivity analysis. The indexing problem is illustrated in Table 4-4. The extreme case of $tp = 3$ and $m = 3$ is considered for a hypothetical hydrograph segment of 11 days. The segment has a decreasing part of 8 days and a recession part of 5 days while excluding the three decreasing days (criteria in Table 4-2). For an h value of 0, all points labeled as recession are predicted. The REF input is simply defined by a translation of tp time steps ahead. Reference states would be decreasing, but not especially recession states, and the embedding M_X may include reconstructed states of non monotonically decreasing sequences. This scenario is potentially problematic at different levels. First, the fact that recession points are forecasted from states under the influence of rainfall is not in phase with the paradigm that recession states are only functions of previous recession states. Second, this effect only occurs for high m and tp, and its magnitude is supposed to be dependent on the recession extraction method constraint about minimal segment length. With $h = 2$, the embedding includes decreasing points exclusively, still, potentially under the influence of rainfall or some interflow components.

Table 4-4: Example of indexing of reference states REF based on the truncation input h illustrated with a prediction horizon tp of 3 and an embedding dimension m of 3.

	time t	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	Q_t	5.0	6.0	7.0	6.7	6.3	6.0	5.7	5.4	5.1	4.9	4.6
Example	is decreasing	-	-	-	1	1	1	1	1	1	1	1
	is recession point	-	-	-	-	-	-	1	1	1	1	1
	is reference point	-	-	-	1	1	1	1	1	-	-	-
$\boldsymbol{h}$**= 0**	is embedded	-	1	1	1	1	1	1	1	-	-	-
	is predicted	-	-	-	-	-	-	1	1	1	1	1
	is reference point	-	-	-	-	-	1	1	1	-	-	-
$\boldsymbol{h}$**= 2**	is embedded	-	-	-	1	1	1	1	1	-	-	-
	is predicted	-	-	-	-	-	-	-	-	1	1	1
	is reference point	-	-	-	-	-	-	-	1	-	-	-
$\boldsymbol{h}$**= 4**	is embedded	-	-	-	-	-	1	1	1	-	-	-
	is predicted	-	-	-	-	-	-	-	-	-	-	1

As long as h rises, recession points are increasingly exclusively predicted from recession points. However, the number of predicted points decreases accordingly, and recession segments with a length of h or less are completely eroded. Furthermore, high h values focus the analysis on particularly long events and, therefore, on potentially extremely low flow values that are not best suited to be studied with a regressor based on local approximations. For these reasons, the inclusion of parameter h in the sensitivity analysis is justified to discuss performance and the traits related to the complexity and nonlinearity of recession dynamics. Input h must be set independently of m and tp, so that h determines the number of points predicted for a given recession method, which means that the same processes are forecasted and interpreted regardless of the selected prediction horizon or dimension. The value of h was limited to 4 because with higher values, the number of predicted recession points was too low.

The Sobol' GSA method requires a number of simulations (input samples) of $N = M\,(2i + 2)$, where M is a replication factor typically between 512 and 2048 and i is the number of input factors considered. In this case, for i= 8 variable inputs a number of N= 36864 simulations were performed at the

University of Florida high-performance computer. The Sobol' sampling and calculation of indexes from the outputs were performed with the SIMLAB v2.2.1 software (Saltelli, 2004).

4.3 Results

4.3.1 Recession Extraction

The statistics resulting from the recession extraction methods (Table 4-2) are compared in Table 4-5. The BRU method leads to the smallest number of points and segments due to its stricter criteria. BRU points and segments are unequally shared between stations compared to the other methods. The BRU method captures flows with a higher decreasing rate. The EDM method also tends to provide short segments. On average, the EDM method captures recession points having the lower mean Q_t and, consequently, the lower recession rate. The KIR method captures the highest number of points and segments of varying lengths that likely include various kinds of flow aside baseflow. The VOG method provides the most extended recession segments.

For each recession extraction method and station in Table 4-5, Figure 4-4 reports the corresponding recession plot of $\log(Q_t)$ against $\log(-dQ_t/dt)$, with $dQ_t/dt = Q_t - Q_{t-1}$ (Eq. 1.6; Table 4-1). In particular, the EDM method extracts recession points yielding a dense and closer recession plot on which patterns are more clearly visible. For comparison with the EDM-Simplex approach, the nonlinear exponent b of the B&N model (Eq. 1.5) is reported using three acknowledged recession fitting methods: the ordinary least squares (Vogel and Kroll, 1992), the lower envelope (Brutsaert, 2008) that involves a quantile regression such that 5% of the recession points are below the regression line, and a standard error weighted regression (Kirchner, 2009). The last one consists of regrouping recession points into bins of various sizes and fits the regression line on the bins' means using weights relative to the bins' inverse standard error (Stoelzle et al., 2013).

Table 4-5: Comparison of recession extraction methods

Extract.	**BRU**			**EDM**			**KIR**			**VOG**		
Station	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Count												
points	57	88	100	192	199	196	798	852	866	224	239	248
segments	33	41	44	79	85	80	177	186	185	24	25	26
Segment length												
min	1	1	1	1	1	1	1	1	1	7	7	7
mean	1.7	2.1	2.3	2.4	2.3	2.5	4.5	4.6	4.7	9.3	9.6	9.5
max	5	9	8	7	7	7	27	30	30	20	22	22
Q_t **[m³.s⁻¹]**												
min	0.58	0.19	0.99	0.39	0.28	1.00	0.35	0.17	0.94	0.47	0.28	0.95
mean	2.65	4.25	5.04	2.24	2.59	3.22	3.10	3.68	5.01	3.00	3.39	4.53
max	7.46	14.1	15.7	11.7	10.9	12.9	11.8	17.8	24.2	7.23	10.7	14.5
$-dQ_t/dt$ **[m³.s⁻¹]**												
min	1.46	3.68	3.09	1.99	2.04	2.31	2.31	5.71	5.45	1.29	2.35	2.69
mean	0.31	0.63	0.64	0.19	0.25	0.26	0.29	0.45	0.55	0.25	0.34	0.40
max	0.03	0.03	0.02	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$

Apart from the weighted method applied to the low number of recession points provided by the BRU extraction method, all combinations agree on a $b(\mathrm{S3}) > b(\mathrm{S1}) > b(\mathrm{S2})$ ranking of nonlinear exponents. Based on this framework, the streamflow series exhibiting the most linear recession is S2. As S2 is a perched section of the stream, one can assume that the B&N model captures the leakage through the permeable river bed, with a Darcian linear-dominant process governing S2 recession. Also, S3 exhibits a higher b exponent. This bias appears mostly due to the minimal flow ensured by the Vauclusian spring (~1 m^3/s) that truncates the points for low flows.

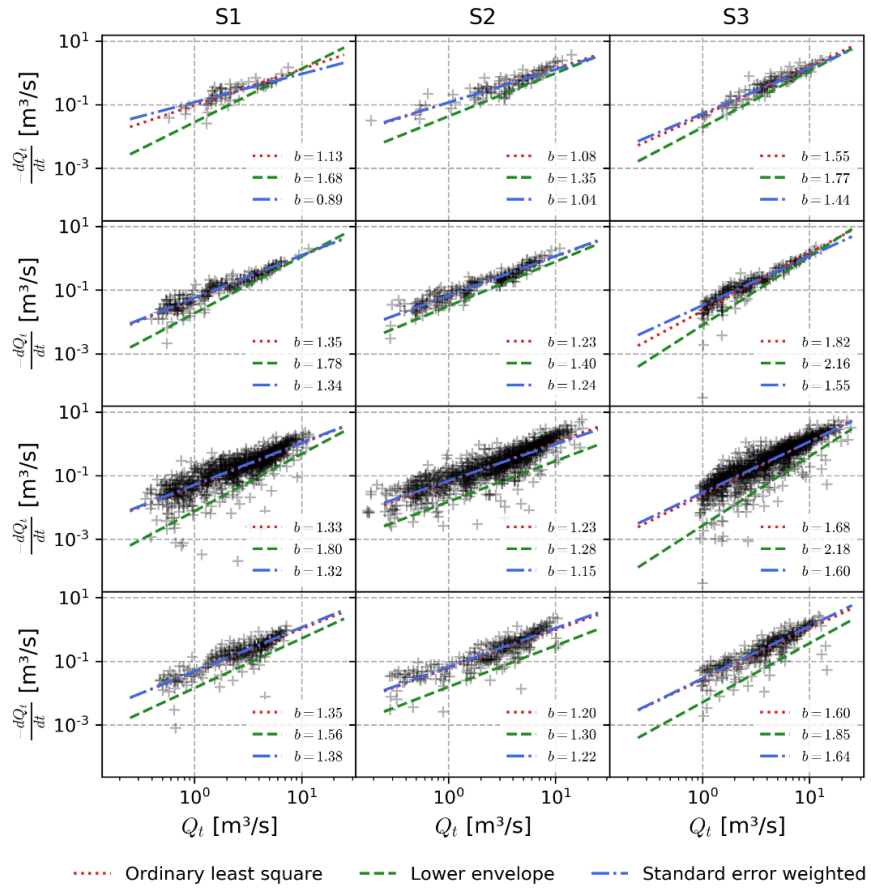

Figure 4-4: Recession plots associated with the recession extraction methods (Table 4-2) for the three stations S1, S2, and S3.The recession point cloud is fitted using the Brutsaert and Nieber (1977) model (Eq. 1.5) using three recession fitting methods: the ordinary least squares (Vogel and Kroll, 1992), the lower envelope (Brutsaert, 2008), and the standard error weighted method (Kirchner, 2009) based on the implementation of Stoelzle et al. (2013). The resulting nonlinear exponents *b* (Eq. 1.5) are reported. All combinations agree on a b(S3) > b(S1) > b(S2) ranking of nonlinear exponents suggesting that S2 is the most linear pattern.

4.3.2 EDM-Simplex GSA

Figure 4-5 presents the result of the Sobol' GSA indices. The most important factors influencing the forecasting performance are REF, h, tp, L, and m by order of importance. The high importance of REF and h suggests that recession forecasting is highly sensitive to the recession extraction method.

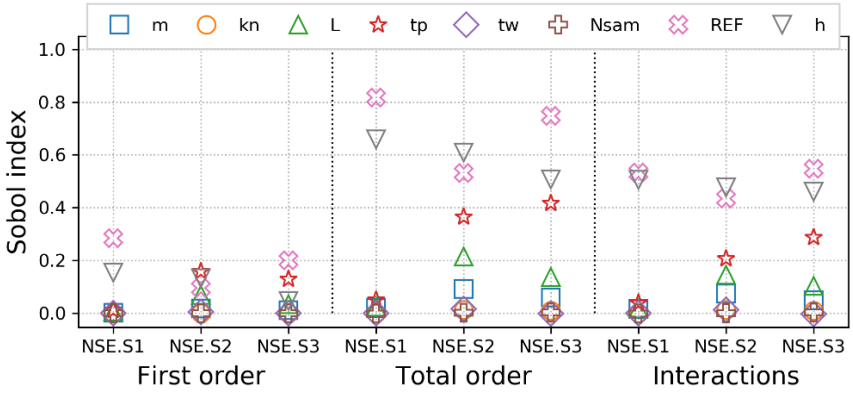

Figure 4-5: First- and total-order indexes of Sobol' sensitivity analysis and their differences.

Globally, the upstream station S1 is found less sensitive to the prediction horizon tp, the embedding dimension m, the size of the library L. Thus, S1 recession dynamics are less complex and less sensitive to initial conditions than S2 and S3. Still, the important differences between the first-order (S_i) and total-order (S_{Ti}) indices suggest that higher-order interactions are occurring between the factors. The sums of the first-order indices $\sum S_i$ are 0.46, 0.48, and 0.42 respectively for S1, S2, and S3. These values stand below the model additivity threshold of $\sum S_i = 0.6$ proposed by (Cukier et al., 1978), suggesting that the EDM-Simplex model output with the studied configuration (Table 4-3) cannot be considered additive and is dominated by input factor interactions. As a result, if the forecasting performance is sensitive to REF and h, the inference about the nonlinear behavior through tp or m is sensitive as well because of the revealed interactions. Therefore, the results below will be reported in relation to the inputs REF and h and for a value of L greater than 200, guaranteeing convergence of the results. Inputs of near-zero importance, k_n, tw, and N_{SAM}, are not discussed as they do not affect the variance of the prediction performance.

The variability of the EDM-Simplex performance for each REF and h combination is represented by the cumulative likelihood of the NSE obtained from the Sobol' simulations (Figure 4-6). In general, the method performance is very good ($NSE > 0.9$) except for the BRU method, especially for h between 2 and 4, and the EDM method with $h = 4$. These cases are those of small numbers of predicted points, respectively 9, 4, and 1 for the most affected case BRU-S1 and 5 for EDM-S2 (Table 4-5). In other cases, the increase in h, especially from low h, has gradually improved forecasting skills. This observation is in line with the idea that the process becomes increasingly deterministic as the recession progresses. The BRU segments, regardless of h, are the least predictable because they are less numerous and recurrent in the hydrograph. The EDM extraction method presents the best forecasting performance, not surprisingly, since the points were extracted using the same model and following the performance logic. However, the forecasts remain accurate regardless of the EDM-Simplex configuration used, most notably regarding the sensitive parameters tp and m. The extraction methods having less strict criteria, namely, VOG and a fortiori KIR, also present excellent forecasting skills. For these methods, S1 appears as being more predictable in general, especially when h is low (i.e., while considering the early part of recession segments). This finding supports the hypothesis that the early stages of the recession, potentially interflows, are more indicative of the geomorphological complexity of the watershed.

Figure 4-7 allows us to study further the particular effects of the prediction horizon tp (Figure 4-7.a) and the embedding dimension m (Figure 4-7.b) for the value of $h = 1$. Regarding tp, all median forecasting skills for a tp of 1 day have an excellent NSE value of >0.99. For each method REF and stations S1, S2, and S3, the results exhibit a statistically significant decrease in the median NSE to a unit increment of tp. Hence, tp is indeed an important factor related to sensitivity to initial conditions, that is, nonlinearity. Between stations, S1 exhibits in general (All REF) a slighter decrease such that the median NSE values for tp higher than 1 day are significantly higher than the corresponding values attached to S2 and S3. This is not the case of BRU, either due to the criterion imposed by the method (Table 4-2) to the higher average recession rate or as an artifact due to the low and unequal number of recession points (Table 4-5). This is not the case of EDM either because S3 has higher median prediction skills and the slightest decrease of the median with respect to tp. Potentially, this pattern can be explained by the fact that the EDM method essentially captures low flows (Table 4-5). In the case of S3, these

flows are mainly under the influence of the Vauclusian resurgence that stabilizes the recession rate (Figure 3-1 and Figure 4-4 - EDM vs. S3), which consequently appears to be less sensitive to initial conditions. With the KIR and VOG methods that extract higher flows (Table 4-5), the general pattern (All REF) holds, with S1 having the less sensitive dynamics to initial conditions.

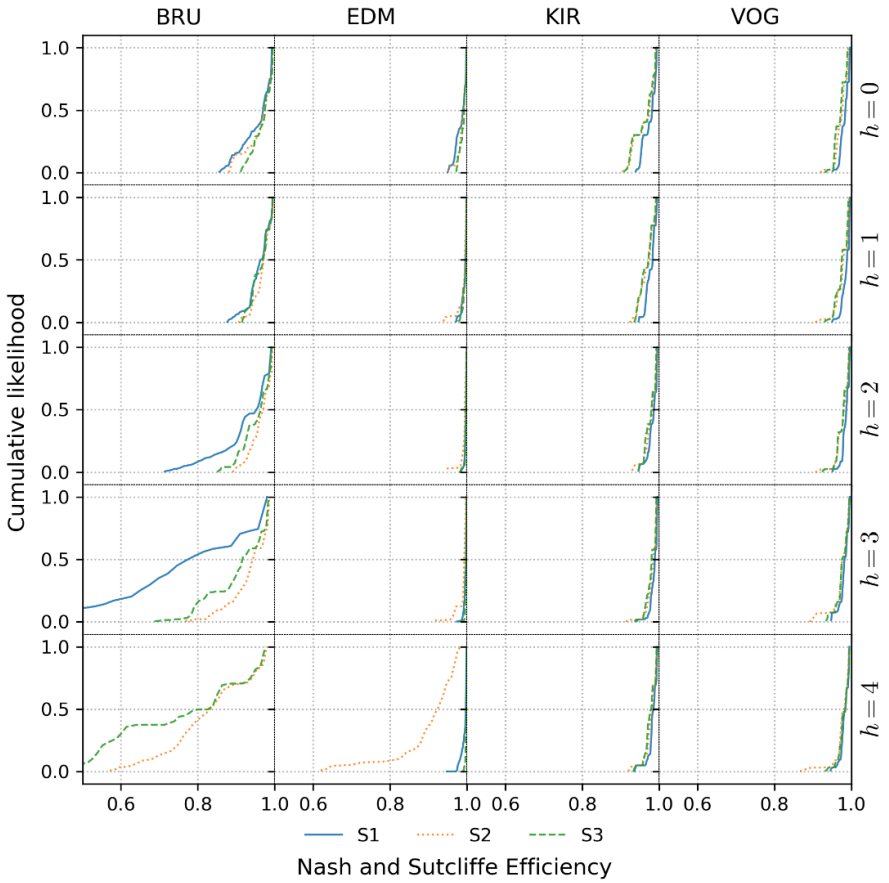

Figure 4-6: Cumulative likelihood of EDM-Simplex Nash and Sutcliffe efficiency (NSE, Eq. 1.4) per recession extraction method and truncation parameter h for the three stations S1, S2, and S3. The number of values for each curve is between $n = 669$ and $n = 712$.The columns refer to the recession extraction methods (Table 4-2), and the rows to the recession segment truncation parameter h (Table 4-3 and Table 4-4). The library length input L is filtered to be higher than 200 to ensure that EDM-Simplex has converged. The S1 cumulative distribution for the BRU recession extraction method and $h = 4$ is missing due to an insufficient number of predictions to calculate NSE.

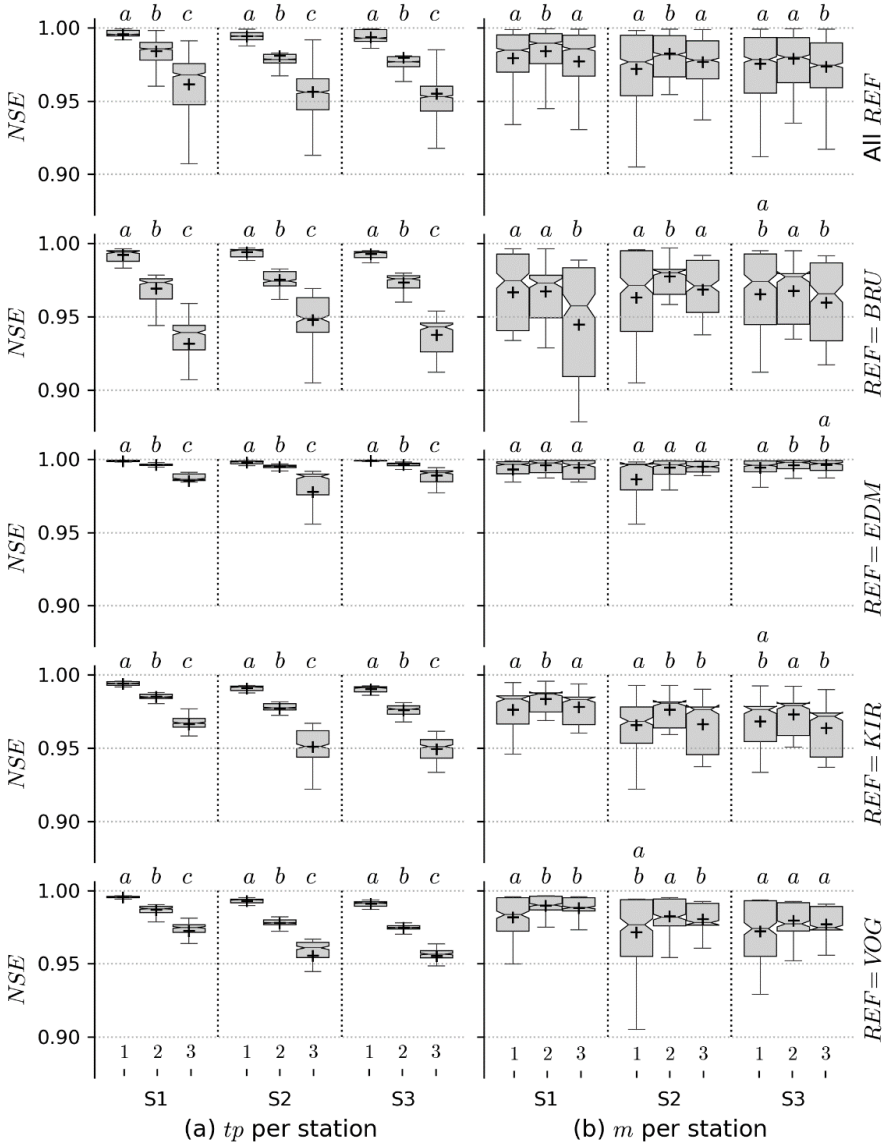

Figure 4-7: Global sensitivity to the (a) prediction horizon and (b) embedding dimension. The boxplot reports the statistics of NSE forecasting performance. The results are presented for L greater than 200 and $h = 1$ (Table 4-3; Table 4-4). The first row, All REF, represents the average behavior of all recession extraction methods of Table 4-2 with the number of samples per box statistics ranging between 889 and 961. Each successive row presents the results for a specific recession extraction method with the number of samples per box statistics ranging between 163 and 307. Letters a, b, and c discriminate groups of statistically significant different medians ($\alpha = 0.05$) (McGill et al., 1978).

The embedding dimension m is less important in explaining NSE's variance, as shown by the larger boxes and their weak separability into distinct groups, a, b, and c, of significantly different medians. This corroborates the GSA results (Figure 4-5). Hence, interpreting m requires caution. Still, $m = 2$ produces better forecasts in general. As shown in Figure 4-5, m is somewhat more influential at station S2. Figure 4-7 confirms it by showing that the S2 forecasting performance is inferior with $m = 1$. In general, S1 has higher predictability with $m = 1$ compared to the other stations, indicating a lesser complexity.

4.4 Discussion and Perspectives

Although EDM-Simplex allies parsimony and some degree of interpretability, at the moment, our empirical approach does not fully replace the physically-based ones relying on parametric equations because sensitivity to initial conditions is not directly interpretable physically. Most likely, substantial advances should be made by (1) hypothesis testing, (2) alternative data crossing, (3) catchment comparison, and (4) model evaluation. These four perspectives are discussed in the next ending sections and are generally associated with an analytical framework for nonlinear time-series analysis, recently applied to hydrological case studies (e.g., Huffaker et al., 2016; Medina et al., 2019). Hopefully, these will allow connecting the EDM framework closer to the operational goal of recession analysis (i.e., the estimation of mass balance components or the study of climate or anthropogenic impact on water resources), beyond the forecasting purpose of this initial work.

4.4.1 On Hypothesis Testing

Based on the results from this work, for an adequate use of the B&N model (i.e., where and when the hypotheses of a Boussinesq's aquifer are applicable), the EDM-Simplex should show high predictability and a low decrease in the forecasting skills over time. Since this predictability and decrease are likely to depend on the recession extraction method, the most critical point in order to use EDM-Simplex consistently, or in general for making advances in recession analysis, is that extraction methods must be able to capture segments of hydrographs that correspond to the baseflow discharged by the aquifer. This becomes an important step to connect the dots between empirical and physically-based recession approaches, as discussed below. EDM-Simplex

can also bring consistency between different empirical approaches. For example, if $m = 1$ yields the best forecasting performance for a particular river, this would indicate that the recession is mostly a function of time and that the empirical characterization of Q_t as a function of time, that is, the master recession curve (e.g., Lamb and Beven, 1997), would also be justified for studying the behavior of the recession in that case.

In addition, proper inference about recession dynamics requires testing for the stationary hypothesis. In theory, EDM-Simplex assumes structural stationarity (section 4.2.1) and may perform poorly in a study case with noticeable climatic or anthropogenic impact on water resources (e.g., Bogaart et al., 2016). In such cases, the inadequacy of EDM-Simplex can be leveraged to assess empirically if a hydrograph series is under changing conditions. Assuming that change implies that the past is no longer a good basis to predict the future and vice versa, constraining the LIB and the REF input such that one maps the past and the other the future, EDM-Simplex should perform significantly worse compared to an unconstrained model, allowing EDM-Simplex to select neighbor states closer in time. Such a nonparametric test of stationarity is analogous to the cross-prediction test suggested by Schreiber (1997).

4.4.2 On the Use of Alternative Data

In our study case, the use of additional data or conceptual knowledge indicative of the catchment hydrogeology greatly helped in the interpretation of the results of the empirical approach. Still, the data were not directly incorporated in the suggested framework that is designed for univariate time-series analysis. Given that the predictability was already high, the use of alternative data, such as rainfall and evapotranspiration data, piezometric levels, and on-site tracing tests, will most beneficially improve the targeting of actual baseflow recession points. If our proposed EDM recession extraction method (Table 4-2, Appendix V) was set up on ancillary criteria, such alternative data could greatly help in configuring EDM-Simplex to retrieve actual recession points compatible with a hydrological definition of baseflow as underground discharge. As the proposed EDM recession extraction method has the advantage of also retrieving short recession segments, by interpolating these short segments, one can estimate the proportion of baseflow in the hydrograph, that is, a baseflow index. Still, as EDM-Simplex is a univariate empirical method, it cannot be used to estimate baseflow under runoff or

interflow conditions. Its multivariate counterpart called Convergent Cross Mapping (CCM) (Sugihara et al., 2012), can be used and adapted to explore empirically and beyond existing rough multivariate interpolation schemes the case of baseflow separation. Another option is to use a multivariate embedding approach (Deyle and Sugihara, 2011) instead of Takens's embedding (Eq. 4.1), for instance, an actual state space of discharge, piezometric levels, and evapotranspiration data. In general, these perspectives should be tested first in a non-karstic watershed to avoid the problem of substantial hidden heterogeneity and the presence of additional flow components routed by the karst conduits. However, if the reconstruction of the state space includes data other than streamflow, special attention should be paid to the issue of noise, as noise interferes with the detection of the nearest-neighboring states. A first step would be to consider smoothed data or seasonal trends in evapotranspiration or groundwater variations.

4.4.3 On Catchment Comparison

In the hypothesis that recession extraction methods offer the opportunity to compare recession dynamics consistently (see section 4.4.2), EDM-Simplex can be introduced in the catchment comparison framework based on its capability to test hypotheses (see section 4.4.1). An interesting aspect would be to explore further the evolution of recession sensitivity to initial conditions across scales. In this study case, it is observed that at a small scale, the increasing size of the catchment yields an increase of recession sensitivity to initial conditions due to the presence of a karst system. However, one cannot hypothesize that the sensitivity increases monotonically with the catchment area. Indeed, complexity never explodes in nature. Thus, it is expected that at a given scale, catchment organization reduces the complexity of the dynamic into a simpler one, that is, exhibiting less prediction decay through time. These potential applications provide an opportunity to use EDM-Simplex in a catchment comparison framework based on nonlinear dynamics concepts (e.g., Sivakumar and Singh, 2012) while keeping the focus on the recession or the decreasing limbs of the hydrograph as a workaround to major challenges when applied to the full streamflow series (Koutsoyiannis, 2006b). However, given the high performance of EDM-Simplex forecasting, the influence of instrumental and environmental noise during the recession on nonlinearity estimation should be further studied.

4.4.4 On Model Evaluation

The EDM-Simplex algorithm was first introduced to distinguish chaotic determinism from randomness. In time-series analysis, models are traditionally evaluated by checking the residual distribution and their linear autocorrelation (Box et al., 2008). Regardless of the model type used to forecast the recession or other important hydrological dynamics, EDM-Simplex can be used to forecast the residuals and check if there remains any low-dimensional deterministic pattern. This could be done by reporting the model performance against increasing sample size L. If an increasing trend up to a significant performance is found, it would indicate that the model used for recession forecasting includes nonlinear time-dependent errors, thus opening a perspective for model improvement even if the residuals pass the usual tests for randomness and the absence of linear autocorrelation (Ljung and Box, 1978). The test can be done using various embedding dimension m, especially if EDM-Simplex is used, considering that one cannot extract dynamic patterns and diagnose the residuals with the same configurations.

4.5 Conclusion

The EDM-Simplex model (Sugihara and May, 1990) was suggested as an empirical and parsimonious way to forecast recession dynamics and infer the nonlinear behavior of the recession from the sensitivity of forecasting performances to initial conditions. The model was tested with GSA on three hydrograph time-series (S1, S2, and S3) of the Lhomme river (Belgium) being respectively located before, inside, and after a karst system. The GSA results showed that the forecasting skills are excellent regarding the 1-day-ahead forecasts. The median NSE was above 0.99 for all time-series, regardless of the recession extraction methods identifying the hydrograph points being forecasted. In virtue of parsimony, the good performance of EDM-Simplex discourages the use of forecasting methods of unnecessary complexity to the case of recession modeling without justification. The GSA further highlighted that the prediction horizon, defined from 1 to 3 days, is the most important factor related to sensitivity to initial conditions. It is expected that the more nonlinear the recession dynamics are, the more the forecasting performance will decrease with the prediction horizon. However, the absolute forecasting skills at a horizon of 2 and 3 days and the decreasing rate in performance were also sensitive to the factors ruling the recession extraction methods. In particular, the results suggested that recession extraction methods that

consider some transient flows or the early stage of recession with less strict criteria on the exclusion of abnormal recession rates allowed to discriminate better the upstream recession processes from the ones being more nonlinear as affected by the presence of the karst system. In contrast, this increase in nonlinearity as the catchment size includes the karstic area was not observed with the B&N (Brutsaert and Nieber, 1977) parametric model on all recession points. The latter showed that station S2 located in the karstic system has the most linear dynamics, potentially, as section S2 of the river is perched and subject to Darcian percolation flows. Given that both models gave different but interpretable outcomes, the complimentary use of parametric methods is not discouraged as long as their parameters are reasoned empirically when and where they should.

In general, the relativity of recession nonlinearity to the recession extraction method is a shared concern in recession analysis (Stoelzle et al., 2013). This variability is a blessing in disguise and should be investigated, especially if the catchment recession dynamics is considered as a multiple reservoirs problem (e.g., Clark et al., 2009; Harman et al., 2009) that could eventually become disconnected from the river channel (Biswal and Marani, 2010). In that matter, empirical approaches are more suited since their scopes extend beyond the ones of the current physically-based parametric approaches, inherently limited by their static physical assumptions on processes and catchment geomorphology. On the other hand, empirical results are difficult to interpret physically without a priori information on catchment hydrogeology. Thus, the confrontation of empirical evidence with physical assumptions will undoubtedly pave the way for future advances in recession analysis to the point where they will eventually meet by relating time-variant nonlinear patterns to catchment hydrological states and corresponding processes. Such progress would also bring the empirical framework closer to recession analysis's operational objectives, beyond the forecasting objective explored and achieved in this article. For this purpose, our perspectives suggest how the EDM-Simplex model or its variants could be involved in this challenge.

At this point, the results support the use of EDM-Simplex as a parsimonious and performant forecasting tool. Secondly, as EDM-Simplex has proven to be robust in measuring the nonlinearity induced by the karst system, the method provides an indicator of recession nonlinearity that better corresponds to the geomorphological and hydrological complexity of the system. This indicator

can guide modelers towards a conceptual model that considers the relative complexity of the system or in a watershed comparison framework to reveal the spatial patterns of complexity and organization at larger scales.

Chapter 5 Time-series clustering approaches for the dimension reduction of a real time-lapse electrical resistivity dataset

"There is a difficulty about the part and the whole [... it] is to know if the whole and the part form unity or plurality, and how they are one or more, and if they are many, how many?"

Aristotle

Foreword

Before the thesis, I had not yet had the opportunity to develop an expertise on ERT datasets (section 3.3.2). It was not the primary objective of MIGRADAKH (section 1.2.1) to develop one or study in-depth the dimensional reduction of an ERT dataset. I valued the dataset for its temporal length and the spatial information it contains, but its large size posed a problem for applying causal inference methods. Naturally, I wanted to group the time-series using a clustering method. At first, I had a feeling that this task would be solved quickly, within a few days, allowing finding out which method to use. While investigating the literature on clustering and post-processing of ERT datasets, I realized that the task was not trivial. The literature proposes a few specific applications rather than methodological and transversal reviews. Therefore, I decided to look into the issue with the collaboration and support of Arnaud Watlet and Olivier Kaufmann (UMons). The first article was submitted in October 2019 to the Journal of Applied Geophysics. In November 2020, the article is accepted and about to be published under the title: "*Time-series clustering approaches for subsurface zonation and hydrofacies detection using a real time-lapse electrical resistivity dataset*".

Regarding causality (Chapter 2), this chapter has in fact found its rightful place in this document. Finding spatiotemporal patterns is mainly related to the formal cause (d8, see Appendix III) because clustering is a task of organization (section 1.1.2.4). Clustering is an empirical approach (section 2.4.3) grouping time-series based on their similarities following the logic of

induction (d13). While clustering, asking about the appropriate number of clusters is the question of finding the right number of material causes (d7), that we call here hydrofacies. This number aims at being right in the sense of parsimony (d18). This number is also a measure of the system's complexity by reflecting its dimension (section 1.1.2.2).

In particular, the first question in the chapter is "*which algorithm should I used?*". This issue is related to the formal cause (d8) that deals with how things associate and organize themselves. It is also related to the mechanisms (d9) of the method and not the system's physical mechanism. Arguably, the method should be consistent with the system's physical mechanisms, but this issue is not studied because our approach is empirical and based on generic clustering tools. The second question is "*should the clusters be contiguous in space, such as geological materials?".* This question is directly related to the physical meaning (d10) of hydrofacies as material causes (d7) and Hume's principle of contiguity (d23). A third question is about the time-series representation of resistivity series prior to the clustering, another way to question the material causes and their meaning (d7,d10). A final question is directly related to the *problem of induction* (section 2.4.3). It is asked, "*how long should be the ERT dataset to have a robust and time-invariant clustering output?*". In philosophical terms, the question is related to the concepts of *actuality*, what is capable of being seen (d16), and *potentiality*, what is capable of being built (d17). The paper is methodological and entirely empirical. It shows the results for various ways of doing clustering.

In general, without constraining clustering with a long enough dataset or other complementary datasets (d16), including physical constraints or taking into account mechanisms (i.e., the efficient causes, d9), or with a more precise definition or practical utility of hydrofacies (d7 thought with d10 in mind), the paper concludes that the potentialities (d17) are too important. There is a risk for scientists to fall into the *problem of induction* and provide non-robust abstractions (d11) with these methods. Too many causal visions relate to complexity in the sense of difficult to understand (section 1.1.2.1, Figure 1-2). It instills a pessimistic feeling of relativism or uncertainty through all the potentialities that are uncovered (section 2.4.4). However, as pointed out in Chapter 2, acknowledging uncertainty as a fact and a property of complex systems is crucial for understanding. It is also the necessary (d20) step paving the way for scientific progress. It highlights the disposition (d28) to be taken for the next applications to limit and constrain uncertainties.

Abstract

One main application of electrical resistivity tomography (ERT) is the non-invasive detection of geological or hydrological structures in the shallow subsurface. This chapter investigates the capability of time-series clustering to retrieve such features on real time-lapse ERT datasets considering three aspects: (1) the comparison between three clustering algorithms k-means, hierarchical agglomerative clustering (HAC), and Gaussian Mixture Model (GMM), including the question of the optimal choice of cluster number and the identification of resistivity series whose classification is uncertain, (2) the effect of adding a spatial constraint in clustering, and (3) the robustness of the approaches to various representations of resistivity values and the number of time-steps involved in the clustering. The real time-lapse ERT dataset is obtained from a profile installed on the top of the Rochefort cave in Belgium. It consists of resistivity time-series defined over 465 days and associated with 1558 cells of the 2D ERT models derived from a time-lapse inversion. The clustering results are appreciated using clustering validation indices and further confronted with the expert-based structural model of the site.

The three clustering algorithms provide similar spatial patterns on the standardized data and reveal correlated resistivity time-series. Some clusters remain spatially split and regroup time-series with a wide range of mean resistivity, suggesting different geological units within these groups. Clustering on the raw resistivity time-series may also appear inconsistent as the averaged resistivity series per cluster are highly correlated, thus missing the hydrological and functional traits of the subsurface elements. On standardized data, applying a constraint to retrieve spatially tied clusters increases the number of clusters. The grouped series are more homogeneous in terms of mean resistivity due to their spatial proximity, but some inconsistencies remain. Applying the clustering to various time-series representations allows gaining confidence about the redundant spatial patterns. However, the patterns obtained from the full standardized dataset cannot be reproduced from continuous sub-samples up to 100 days, but well from less than 20 samples picked randomly over the 465 days. These results show the critical impact of serial correlations in the clustering process and suggest to monitor the surface systems in a wide range of environmental conditions. Accordingly, our study highlights the importance of time-variable parameters in identifying structural facies and hydrofacies with ERT while demonstrating the strength of long-term monitoring.

5.1 Introduction

The electrical resistivity of surface soil varies with the mineralogical composition of soil and rocks, temperature, the water content, and its solute composition. Electrical resistivity tomography (ERT) is a technique commonly used in geosciences that aims to capture these variations. ERT relies on electrodes, a current injection scheme, and the inversion of an associated resistivity model to map the resistivity of the shallow subsurface, either in two or three dimensions, and derive geological and hydrological interpretations (Banton et al., 1997; Samouëlian et al., 2005). Time-lapse ERT extends to an additional dimension by using repeated current injections over time, allowing the retrieval of temporal variation of resistivities. For their capabilities of generating a large amount of spatialized data at low cost, time-lapse ERT has been increasingly used in the near-surface geophysics community to investigate subsurface geology or hydrogeological processes (Barker and Moore, 1998; Kuras et al., 2009; Singha et al., 2015).

The visual or computer-assisted interpretation of inverted resistivity models remains challenging as the inversion procedure often relies on smoothness constraints, producing fuzzy patterns rather than a clear representation of subsoil heterogeneities (e.g., Günther et al., 2006; Loke and Barker, 1996). Besides, the resolution decreases, or the uncertainties of the inversion image increase as a function of the distance to the electrodes (e.g., Hermans and Irving, 2017). Accordingly, the improvement of the ERT models is approached from different angles. The first one focuses on the inversion procedure itself, for instance, by considering adaptable constraints to produce sharper results (Fiandaca et al., 2015; Nguyen et al., 2016). A second option, detailed in the next paragraph, is to apply post-inversion processing to enhance the interpretability of the model outputs. Other refinements may come from crossing strategies and datasets: the joint inversion of multivariate geophysical data (Doetsch et al., 2010; Infante et al., 2010) or the definition of an ensemble model either from a distribution of inversion parameters (Audebert et al., 2014) or from multiple electrode configurations (Ishola et al., 2015). Similarly to Paasche and Tronicke (2007), post-inversion approaches can be coupled with the inversion strategies by an iterative procedure (Doetsch et al., 2010; Elwaseif and Slater, 2012; Infante et al., 2010; Singh et al., 2018; Zhou et al., 2014) and be part in a fully automated way of an integrated ERT monitoring and modeling environment (e.g., Wilkinson et al., 2019).

In particular, post-inversion approaches can be defined in a mutually non-exclusive way, according to different aspects. Some papers target the detection and zonation of static features such as geological boundaries or structures (Caterina et al., 2013; Chambers et al., 2012, 2013; Doetsch et al., 2010; Hsu et al., 2010; Kutbay and Hardalaç, 2017; de Pasquale et al., 2019; Xu et al., 2017), defects in covered landfill (Genelle et al., 2012), buried archeological objects or cavities (Elwaseif and Slater, 2010, 2012). Feature detection can also be improved on multivariate models or datasets (Di Giuseppe et al., 2014, 2018; see also Paasche et al., 2006). Other applications cover dynamic processes: mapping of the water or leachate infiltration front (Audebert et al., 2014; Scaini et al., 2017), the tracking of tracer's motion (Ward et al., 2016), or groundwater level monitoring (Chambers et al., 2015).

In any of these applications, the underlying algorithms can be summarized into the three following types: (1) gradient edge detection, (2) object segmentation into two groups (binarization), or more through (3) unsupervised classification, i.e., clustering. Clustering algorithms have some merits compared to other techniques. Ward et al. (2014) mentioned that gradient edge detection methods are limited since the steepest gradients are not always concurrent with geological interfaces, especially given the smoothness-constrained inversion and the lack of resolution at depth in ERT images. Gradient edge detection is also applied in contexts where the substrate is typically organized in successive horizontal layers. As such, the applicability of this algorithm is challenged for anisotropic heterogeneous environments such as karst systems. Compared to segmentation algorithms that divide models into two subgroups, clustering algorithms have the advantage of not limiting the number of groups that can be defined according to their distinct resistive behavior, which on the other hand, raises the problem of the optimal choice of the number of clusters.

Overall, a few distinct clustering algorithms have been applied: fuzzy c-means (Chambers et al., 2015; Kutbay and Hardalaç, 2017; Paasche et al., 2006; Paasche and Tronicke, 2007; Singh et al., 2018; Ward et al., 2014), k-means (Audebert et al., 2014; Di Giuseppe et al., 2014, 2018; Ishola et al., 2015; Scaini et al., 2017), Gaussian Mixture Models, GMM (Doetsch et al., 2010), and Hierarchical Agglomerative Clustering, HAC (Genelle et al., 2012; Xu et al., 2017). The fuzzy c-means, k-means, and GMM algorithms belong to the family of iterative relocation clustering algorithms. The fuzzy c-means and GMM are similar because they yield to probabilistic clusters, also termed soft

or fuzzy clusters, meaning that an item may be assigned to several clusters with a given probability. On the contrary, k-means is a hard or crisp clustering algorithm according to which each item is assigned to a single group. HAC is another hard clustering algorithm based on a nested structure represented by a dendrogram.

Notwithstanding the availability of these advanced clustering techniques, methodological issues remain when applying clustering to real-world ERT datasets. Fuzzy c-means, for instance, was applied as a fuzzy algorithm to deal with the uncertainties brought by the smoothness of non-time-lapse ERT models (Ward et al., 2014). It is, however, not clear how such algorithms would work with time-lapse datasets. Further, Genelle et al. (2012) and Xu et al. (2017) used the HAC method to cluster for the first time ERT time-series of a time-lapse 2D dataset made of respectively 6 and 20 time-steps. Nevertheless, their study did not address critical issues such as the impact of alternative clustering algorithms on clustering results, the selection of the optimal number of clusters, or the evaluation of either the robustness or the uncertainties in clustering results. Also, Ward et al. (2014) suggested considering the local neighborhood and spatial constraints in clustering processes, an issue, which still needs to be further analyzed.

The chapter focuses on the post-inversion clustering of ERT time-lapse datasets to extract and delineate spatially homogeneous features based on their resistivity patterns and address the above-mentioned concerns. The term hydrofacies, as discussed here, denotes spatial zones of similar patterns in their mean inverted resistivity, standard deviation, and correlation, assuming that they encompass common lithology and synchronous hydrological response at a daily time resolution. In particular, the chapter covers (1) the comparison and parametrization of three candidate clustering algorithms (k-means, GMM, and HAC) while addressing the question of the optimal number of clusters and the evaluation of the clustering results, and (2) the pertinence of including spatially explicit information in the clustering task. Finally, we discuss (3) the robustness of the clustering outputs to various representations of the resistivity data, whether or not log-scaled, normalized, differenced, decomposed, as well as the impact of the time span of the ERT model on the clustering outputs. The analysis is based on a 2D-ERT dataset collected over a 465 days time-domain (section 3.3.2, Figure 3-4) at the RCL study site (section 3.2, Figure 3-2). To further encourage reproducibility and reusability,

programming aspects exclusively relies on Scikit-learn (Pedregosa et al., 2011), an open-source Python package for machine learning.

5.2 Theory and Methods

5.2.1 Time-Series Clustering (TSC)

Clustering consists of grouping high dimensional data into fewer classes based on groups' inner similarities and groups' outer dissimilarities. In particular, time-series clustering (TSC) aims at grouping individual time-series together (Liao, 2005). TSC can be challenging due to the high dimensionality of time datasets: (M, N) where M is the number of time-series (samples) and N the number of time steps (features). Averaging clusters reduces the dimensionality to (k, N) where k is the final number of clusters. A clustering algorithm defines clusters and their members based on criteria involving distance or similarity measures. In machine learning, clustering is also defined as unsupervised classification since there are no predefined labeled groups that could serve as a basis for training.

Due to the combined effect of data structure and dimensionality, the diversity of fields of application, the different clustering purposes, and the nature of hunted patterns, a wide variety of TSC approaches are found in the literature (Aghabozorgi et al., 2015; Liao, 2005). These are, for the most part, declined under three aspects:

1. a time-series representation, which denotes any transformation of the time-series reducing the dimension of the dataset before the clustering;
2. a clustering algorithm relying on a distance measure;
3. an evaluation technique.

A prior reduction of the dataset dimensionality has several advantages: diminution of the memory consumption and speed-up of the clustering algorithm, noise reduction, and the harmonization of time-series data of unequal length or resolution into a dataset where an equal number of features characterize each sample time-series. If no prior reduction is applied, we refer to the raw-data-based approach (Liao, 2005). As far as TSC is concerned in this study, ERT series are univariate, real-valued, uniformly sampled, and smoothed (due to inversion smoothing constraint), of equal length, and relatively short. For these reasons, raw-data-based approaches are considered

until section 5.3.3, where dimension reduction is tested. Still, raw TSC usually involves a scale transformation. The z-standardization is used in the vast majority of cases and applied to each of the M samples, i.e., the individual time-series X_i :

$$X_{z,i} = \frac{X_i - \mu(X_i)}{\sigma(X_i)},\ i\ \in [1, M] \qquad Eq.5.1$$

where $\mu(X_i)$ and $\sigma(X_i)$ stand for the mean and standard deviation estimates for X_i.

Then, in section 5.3.3, four time-series representations, each of them either applied to the resistivity or the log-resistivity, are used: the raw data without transformation, the z-standardized data (Eq. 5.1), the differenced data $(X_i(t) - X_i(t-1))$ followed by z-standardization (Eq.5.1), and the decomposed data using principal component analysis (PCA, see Figure 3-4, section 3.3.2) on the z-standardized data (Eq. 5.1). The prior z-standardization of data or other scaling methods are very common while performing TSC. However, the scaling may not be necessary in this case since time-series are all representative of the same physical variable, i.e., resistivity. Differencing removes the seasonal variation of the mean resistivity and will most likely result in a clustering that is more sensitive to the synchronous response of daily variations. Removing seasonality is motivated as it may dominate the signal and produces a cluster based on their common seasonal patterns while being non-correlated in other frequencies. However, one cannot exclusively consider differentiated data because it can be assumed that a hydrofacies must group series with similar seasonal patterns. By decomposing the covariance matrix of the dataset, PCA reduces its dimension from *N* to several orthogonal components that explain most of the variance of the dataset (Figure 3-4, e to h). Since PCA is applied to the z-standardized data, the covariance matrix is equivalent to the correlation matrix, and the PCA reveals correlation patterns in the time-series across space while removing the noise that may influence the clustering.

5.2.2 Clustering Algorithms

There are no strict restrictions on the use of conventional clustering algorithms for the specific case of TSC. However, it is common to have distance functions modified according to the purpose of clustering. Two cases arise depending

on whether the aim is to group synchronous and linearly correlated series (similarity in time) or whether the procedure must rely on elastic measures of distance tolerant to some distortions or asynchronies (similarity in shape). This paper focuses on the first case, i.e., the similarity in time compliant with our definition of hydrofacies. The similarity in time relies usually on using Euclidean distances, squared Euclidean distances, or correlation-based distance. On a z-standardized dataset (Eq. 5.1), the correlation coefficient $R_{X_i X_j}$ between two time-series is related to their squared Euclidean distance $d^2_{X_{z,i} X_{z,j}}$ such that $R_{X_i X_j} = 1 - d^2_{X_{z,i} X_{z,j}}/2N$. Despite the introduction of new distance metrics, Euclidean-based distances remain the simplest and one of the most competitive options (Keogh and Kasetty, 2003).

An extensive and non-exclusive taxonomy describes clustering algorithms (Tan et al., 2019). An important distinction is based on the clustering structure. If the algorithm produces an independent partition of k clusters, it is called a partitional algorithm. On the other hand, if clustering produces a tangled structure of groups and subgroups, it is referred to as hierarchical, although it is possible to retrieve a partition of k clusters based on a cut-off distance. Another dichotomy is based on the hard (or crisp) or probabilistic (or soft, fuzzy) nature of the partition. Hard clustering labels each object i to one unique cluster, while probabilistic clustering defines a probability of membership. Another type of clustering algorithms is prototype-based or center-based clustering. These algorithms partition objects based on their distance from the centroid of the cluster and, for these reasons, tend to produce convex clusters centered on the mean. In the 2D example of Figure 5-1, clusters A and C are convex since they could be averaged to a characteristic element, the centroid that belongs to the cluster. Cluster A has a spherical covariance matrix, while C has an anisotropic covariance matrix. On the reverse, cluster B is concave, and the centroid is no longer a reliable prototype. In general, concave clusters are extracted using methods that consider the local neighborhoods and densities around each sample (e.g., section 5.2.2.2). It allows extracting dense clusters regardless of their structural arrangement. However, these approaches are challenged in the case of an ERT model given smoothness constraints and the subsequent lack of sharp variations in resistivity.

This study relies on three prototype-based clustering algorithms so that the resistivity series can be averaged into a mean representative series per cluster.

These are k-means, hierarchical agglomerative clustering (HAC), and Gaussian Mixture Models (GMM). The k-means algorithm is partitional, hard, and tends to produce convex spherical clusters. HAC is hierarchical and hard. The covariance structure of the clusters depends on the distance metric and the constraints applied to the agglomeration. Finally, GMM is partitional, probabilistic, and not tight to a spherical covariance. The Python Scikit-learn library (Pedregosa et al., 2011) provides all the clustering algorithm used in this study.

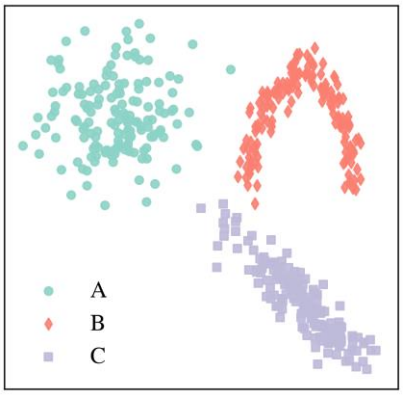

Figure 5-1: Example of three different cluster distributions in two dimensions. Cluster A is convex and spherical. Cluster B is concave. Cluster C is convex and anisotropic.

5.2.2.1 K-Means

The k-means algorithm is the most used clustering method (Berkhin, 2006). It is a partitioning relocation clustering algorithm based on the principle of finding a partition C of k clusters by minimizing the sum of squared Euclidean distances between each object i belonging to a cluster $c \in C$ with respect to the cluster centroid μ_c. The objective function to minimize is then:

$$\phi = \sum_{i \in M} \min_{c \in C} \| i - \mu_c \|^2 \qquad Eq.5.2$$

The original k-means algorithm is referred to as Lloyd's algorithm and consists of a simple series of repeated steps:

1. k clusters centers μ_c are randomly sampled given a uniform probability;
2. Each object i is assigned to the cluster closest center;
3. ϕ is computed with respect to μ_c;
4. A new μ_c is obtained by averaging cluster members.

Steps 2 to 4 are repeated until ϕ is stable. The k-means algorithm tends to produce convex clusters of equal variances across the feature space, i.e., spherical clusters as cluster A in Figure 5-1. The cluster prototype is the centroid μ_c. Due to the random initialization of cluster centers (step 1), a few repetitions of the full process (steps 1 to 4) are usually required to avoid convergence to suboptimal results. The best clustering partition, i.e., with minimal ϕ, is kept. Depending on the dataset, k-means may remain unstable and yield non-deterministic outputs. This study relies on Scikit-learn's implementation of the k-means++ algorithm (Arthur and Vassilvitskii, 2007). The k-means++ implementation improves the speed and accuracy of the original k-means by modifying the randomized initialization scheme (step 1). The idea is to spread initial centers allocation. The first center is still sampled given uniform probability distribution, while the subsequent centers are sampled given probability densities inversely proportional to the distance to previously defined centers.

5.2.2.2 Hierarchical Agglomerative Clustering (HAC)

Hierarchical Agglomerative Clustering (HAC) differs from k-means as it provides a nested structure of the clustering through a dendrogram. HAC uses a bottom-up approach: it starts from individual samples i and merges them into branches based on their proximity until one cluster remains. Clusters are progressively merged based on their relative proximity. The proximity is defined by linkage methods defining the distance between clusters. This study focuses on the Ward linkage method (Ward, 1963) that minimizes the sum of squared differences within all clusters, likewise Eq. 5.2. Hence, the output is quite similar to that of k-means and tends to produce convex clusters of equal covariances (Figure 5-1, cluster A), which could be averaged into a cluster prototype, i.e., the centroid. Unlike k-means, which relies on random

initializations of cluster centers, HAC's outputs are stable and do not require several iterations.

A particularity of the Scikit-learn's implementation lies in the opportunity of constraining the merging of branches by providing a connectivity matrix (Abraham et al., 2014). Such a matrix is binary of square shape (M, M), where M is the number of samples and distinguishes connected objects from disconnected objects so that two branches can be merged only if spatially connected objects exist between them. This functionality could be used to retrieve non-convex or non-spherical clusters in the N-dimensional feature space such as Cluster B and C in Figure 5-1. Usually, the connectivity matrix is computed using a nearest-neighbor approach. In this case, the connectivity matrix is built from the mesh of the ERT model. Two cells are connected if they share an edge. This capability is used in section 5.3.2 as a spatial constraint in order to retrieve spatially homogeneous clusters.

5.2.2.3 Gaussian Mixture Model (GMM)

Gaussian Mixture Models (GMMs) aim at modeling a dataset as a linear mixture of k Gaussian distributions defined in the N-dimensional feature space (Berkhin, 2006). Multivariate Gaussian models are related to the Mahalanobis distance that evaluates samples' distance to a given distribution (Gallego et al., 2013). As a probabilistic algorithm, the clustering is soft so that each object i has a probability of belonging to each cluster. For a given object, these probabilities sum up to one. GMM requires as input the number of clusters k and relies on the expectation-maximization algorithm (Dempster et al., 1977) to find an optimal clustering. Expectation-maximization is closely related to the k-means algorithm as it involves iterative relocations: the starting point is a random initialization of k Gaussian distributions that are iteratively reallocated by updating the GMM parameters, i.e., the mixture k weights, the k mean vector of dimension N, and the k NxN covariance matrix. Doing so, GMM maximizes the overall likelihood L that each object belongs to the Gaussian mixture.

By default, the Scikit-learn implementation of GMM uses the same initialization strategy as k-means++ (section 5.2.2.1) and automatically assign each sample to the most likely group. GMM is non-deterministic and different realizations may give different outcomes due to the random

initialization. Different types of covariance matrix exist. Choosing a spherical type will add the constraint that the variance in each of the N dimensions should be approximately equal. As a result, GMM would yield probabilistic convex spherical clusters, similarly to what would be expected from the k-means and Ward-HAC methods. In this context, GMM is used with no constraints on the covariance so that each cluster may have its specific covariance matrix, and GMM may retrieve anisotropic convex clusters such as Cluster C in Figure 5-1.

5.2.3 Clustering Evaluation

The evaluation of clustering is difficult as it is an unsupervised classification meaning that ground-truth labels are usually not available. However, the literature suggests different clustering validation indices aiming at providing both (1) a statistical evaluation of the clusters to measure how well their members are tight and separated from the other clusters and (2) comparing two different partitions in terms of similarity. Both kinds of indices are used in this study. They are all implemented within the Scikit-learn library.

5.2.3.1 Silhouette Index (SI)

Silhouettes were introduced to measure how well an object belongs to its own cluster and as a tool to objectify the choice of the number of clusters k in partitioning algorithms such as k-means (Rousseeuw, 1987). For an object i part of the M samples, a Silhouette value $S(i)$ relies on the mean intra-cluster distance a_i and the mean nearest-cluster distance b_i:

$$S(i) = \frac{b_i - a_i}{\max\{a_i,\ b_i\}},\ i \in [1,\ M] \qquad Eq.5.3$$

$S(i)$ ranges from -1 to 1 and renders the degree of membership of the object to its cluster. A negative value suggests that the object is assigned to the wrong cluster and stand for an outlier. By averaging Silhouette values, an overall Silhouette index (SI) can be computed to render the clustering quality:

$$SI = \frac{1}{M}\sum_{i} S\,(i),\ i \in [1,\ M] \qquad Eq.5.4$$

Typically, k is chosen in such a way that SI is maximum. A comparative analysis of 30 validation indices reports SI as the best index on various synthetic datasets and in the top tier on real datasets (Arbelaitz et al., 2013). However, SI tends to endorse clustering that produces convex clusters, such as k-means or HAC, and may be inappropriate if the algorithm allows the retrieval of anisotropic or concave clusters. This case may happen with GMM or the HAC algorithm if constrained with a connectivity matrix.

5.2.3.2 Information Criteria

The number of components using GMMs are usually not optimized using the Silhouette index but based on information criteria relying on the log-likelihood of the GMM and accounting for the number of free parameters in the model. For this purpose, the Akaike Information Criterion AIC (Akaike, 1974) and the Bayesian Information Criterion BIC (Schwarz, 1978) are usual:

$$AIC = -2\log(L) + 2d \qquad Eq.5.5$$

$$BIC = -2\log(L) + d\log(M) \qquad Eq.5.6$$

where $\log(L)$ is the log-likelihood, M the number of samples and d is the number of degrees of freedom related to the model. The degrees of freedom d is given by summing the covariance, mean, and mixing weights free parameters. With no constraint on the covariance, the number of covariance free parameters are given by the half of the number off-diagonal elements and the number of diagonal elements, i.e., $kN(N+1)/2$, where N is the number of time-steps in the case of TSC. The number of mean parameters is given by kN since the mean vector is of dimension N. At last, the number of weight parameters is given by $k-1$ since $k-1$ parameters are sufficient to describe the mixture weights as they sum up to one. Unlike SI, AIC or BIC should be minimal for an optimal k.

5.2.3.3 Adjusted Mutual Information (AMI)

Based on the information theory, the adjusted mutual information (AMI) is used to measure the similarity between two partitions, or the classification performance if one partition is considered as ground truth data. Another application is consensus clustering, which aims at identifying a more robust partition from an ensemble of different clustering algorithms' outputs based

on their degree of agreement (Monti et al., 2003; Vinh and Epps, 2009). In this chapter, AMI is used to compare the similarity of two clustering partitions with and without spatial connectivity constraint (section 5.3.2), the outcomes of different time-series representations, or the clustering's robustness derived from subsamples instead from the whole dataset (section 5.3.3). AMI is an adjusted measure of similarity. Adjustment in clustering comparison is needed to account for the expected similarity score of randomness, which may vary according to the number of clusters k. It allows having a similarity score ranging from 0 to 1, with 0 corresponding to the score of random labeling and 1 reflecting a perfect agreement between two clustering outputs. Scikit-learn's implementation of AMI relies on Vinh et al. (2010). Considering two clustering partitions vector **U** and **V**:

$$AMI(\mathbf{U},\mathbf{V}) = \frac{I(\mathbf{U},\mathbf{V}) - \mathbb{E}\{I(\mathbf{U},\mathbf{V})\}}{\max\{H(\mathbf{U}),H(\mathbf{V}))\} - \mathbb{E}\{I(\mathbf{U},\mathbf{V})\}} \qquad Eq.5.7$$

where $H(\mathbf{U})$ and $H(\mathbf{V})$ are the information entropy of the given partition, and $I(\mathbf{U},\mathbf{V})$ is the mutual information between both partitions. The expected mutual information for randomness is $\mathbb{E}\{I(\mathbf{U},\mathbf{V})\}$ based on random partitions preserving the number of clusters k and the number of members in each cluster. In general, AMI has the advantage that its score remains unchanged in case of permutations of the cluster labels. It is particularly useful for comparing agreement between two partitions since one object may belong to two respective clusters that are similar but most likely labeled differently.

5.3 Results

5.3.1 Comparison of Clustering Algorithms

This first section compares k-means, HAC with Ward's linkage method, and GMM. The clustering algorithms are applied to the z-standardized (Eq. 5.1) resistivity time-series (Figure 3-4, d). The appropriate number of clusters k is studied by relying on the Silhouette Index (SI, Eq. 5.4) for k-means, HAC, and GMM. Regarding GMM, the optimal number of clusters is also appreciated using the AIC and the BIC criteria (Eq. 5.5 and 5.6). The results are reported in Figure 5-2. The higher SI yield to the preferred k. On the contrary, the lowers AIC or BIC indicate the preferred k for GMM. Since GMM and k-means presents a risk of non-deterministic outputs due to random initialization, their related curves are represented with error bars relative to 2

standard deviations resulting from 20 runs of the clustering algorithm. Each run involves 20 random initializations for both algorithms to select the best model (see sections 5.2.2.1 and 5.2.2.3). As suggested by the small error bars, the k-means clustering appears stable. In contrast, the GMM model is relatively unstable for k=6 and above with respect to the SI value. This is not the case regarding the AIC or BIC. Hence, the GMM is likely to generate different patterns, although they have a similar log-likelihood L.

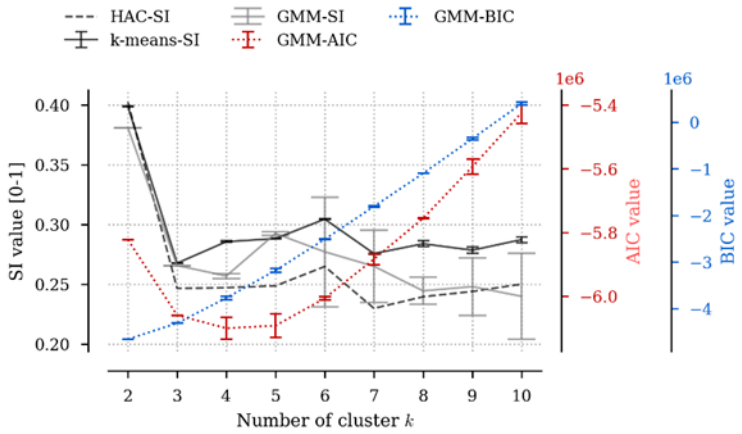

Figure 5-2: Comparison of clustering validation indices for different numbers of clusters and clustering algorithms. The Silhouette index (SI) is reported for the k-means, HAC, and GMM models in black or grey with respect to the left axis. The blue and red lines report on both the right axes the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) obtained for the GMM models. Regarding k-means and GMM, the error bars represent 2 standard deviations across 20 runs, each of them including 20 random initializations.

In general, SI values are relatively low (<0.4), indicating weak compactness and low separability, as expected from a smooth dataset. Still, all indices agree that the optimal number of clusters k is 2, except GMM-AIC (red), suggesting k between 3 and 5. As a second-best, a k value of 6 appears for the HAC and the k-means method, which may be geologically relevant given the different lithologies described in Figure 3-5. The k-means algorithm is stable since almost no deviation in the SI is observed. For comparison, the clustering for k values of 2, 4, and 6 are visualized spatially in Figure 5-3. Cells with white edges are those having a negative Silhouette value ($S(i)$, Eq. 5.3). Since GMM is unstable (Figure 5-2, GMM-SI), the spatial patterns represented in the figure

are the product of one single realization and is subject to changes across runs. This is particularly the case of GMM with k=6 (Figure 5-3.i), for which the displayed patterns were intentionally selected to depict a pattern that differs from the one retrieved by k-means and HAC (Figure 5-3, g and h).

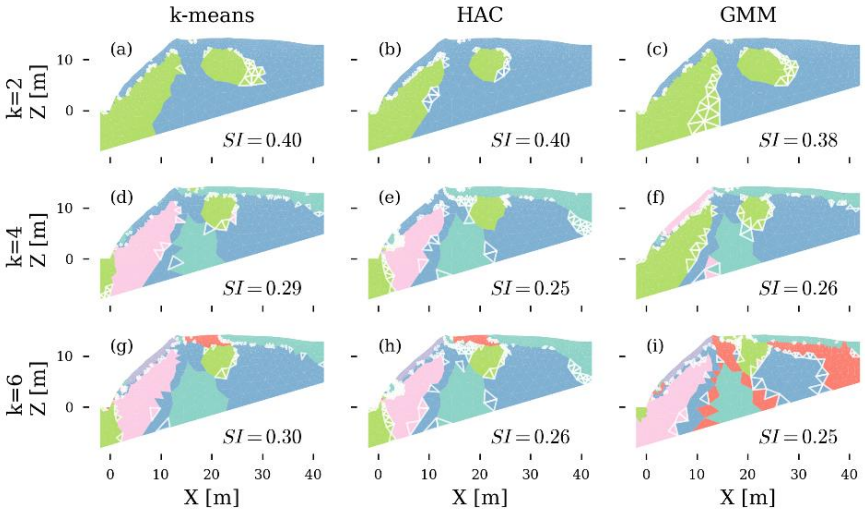

Figure 5-3: Comparison of the spatial clustering patterns for k=2, 4, and 6. The figure columns refer to the three clustering algorithms applied to the time-lapse ERT dataset: k-means, HAC, and GMM. The rows represent different choices regarding the number of cluster k. SI is the corresponding average Silhouette index associated with the partition. The white edges correspond to negative $S(i)$ values (Eq. 5.3) indicating potentially misclassified cells.

Even if some group attribution may differ with GMM (e.g., Figure 5-3.f), the spatial patterns of zonation are generally similar regardless of the method. In general, the patterns of Figure 5-3 match well with the one highlighted by the PCA decomposition (Figure 3-4, f to h). With k = 2, the green cluster is representative of the slope's subsurface (mostly clayey limestone, Figure 3-5, zone G) plus an additional inclusion below the plateau matching roughly the dense limestone group F in Figure 3-5. The green cluster is divided into two parts, pink and green, once k = 4 (Figure 5-3, d and e), except with GMM that instead identified the top part of the slope (Figure 5-3.f, Figure 3-5, zone C). Another split occurs with the corresponding blue cluster. The top surface appears as being dynamically related to the deeper low resistivity area in the fractured zone (see Figure 3-4.a and Figure 3-5). With k=6, the slope's surface appears as a cluster on its own (violet) in all cases. As an additional

comparison with Figure 3-5, the k-means and HAC outputs (Figure 5-3, g and h) present a horizontal division of the plateau into two clusters. The identified red cluster suggests that different dynamics occur at the surface of the fractured area and above the dense limestone area (Figure 3-5, zone F), as visible on the PCA first and third components (Figure 3-4, f and h). The red cluster of the GMM clustering (Figure 5-3.i) is more in phase with Figure 3-5 as it separates the soil surface from the underlying bedrock. In the next sections, HAC will be exclusively considered because it is similar but computationally faster than k-means and does not have stability issues such as GMM.

In contrast with Figure 5-3, Figure 5-4 reports the k=6 HAC clustering applied on the log-resistivity inverted data: the z-standardized log-resistivity for the first column (Figure 5-4.a.x) and the raw log-resistivity for the second one (Figure 5-4.b.x). Applying HAC on the z-standardized resistivity (Figure 5-3.h) or the z-standardized log-resistivity (Figure 5-4.a.1) provides spatially similar clusters on this long term dataset. It means that the log-transformation does not alter much the correlation between time-series. Figure 5-4.a.3 reports the averaged raw log-resistivity time-series. They have distinct dynamical patterns, especially for the delay and magnitude of resistivity declines that occur during fall. The pink and the green cluster (Figure 5-4.a.3) are less responsive to variation in resistivity over time and were regrouped together in the k=2 partition in Figure 5-3. Figure 5-4.a.2 shows that, however, some clusters (e.g., lime green, blue, or turquoise) are spread over the entire statistical space defined by the mean log-resistivity and its standard deviation. Hence, if such a cluster gathers correlated series, it most likely groups different geological materials together, thus, different hydrofacies, which encourage to consider raw log-resistivity as well for more consistency.

Nevertheless, clustering on the raw log-resistivity alone yields to the quantization of the ERT models into iso log-resistivity clusters. This is most visible on the statistical scatterplot of Figure 5-4.b.2. In other words, the clustering of the full dataset of 465 days is roughly equivalent to the clustering of the mean of the 1558 log-resistivity series. The profile is quantized following the order of magnitude of the average resistivity. No information about the dynamical nature of resistivity is leveraged to define the clusters. Consequently, the clustering produces averaged time-series (Figure 5-4.b.3) that are highly correlated, hence, poorly representative of the subsurface system's hydrological states. However, some spatial zones are of interest, such

as the spatial red node within the turquoise cluster below the plateau that map to the porous limestone area of lower resistivity compared to the surrounding denser limestone (Figure 3-4.a, Figure 3-5., zone E). Besides, if some spatially organized patterns are consistent in both approaches, this is not the case of the pink cluster of Figure 5-4.a.1 that should instead be eventually broken up into a lower and upper part (Figure 5-4.b.1). Indeed, while spatially tied, the pink cluster presents a wide range of mean log-resistivity.

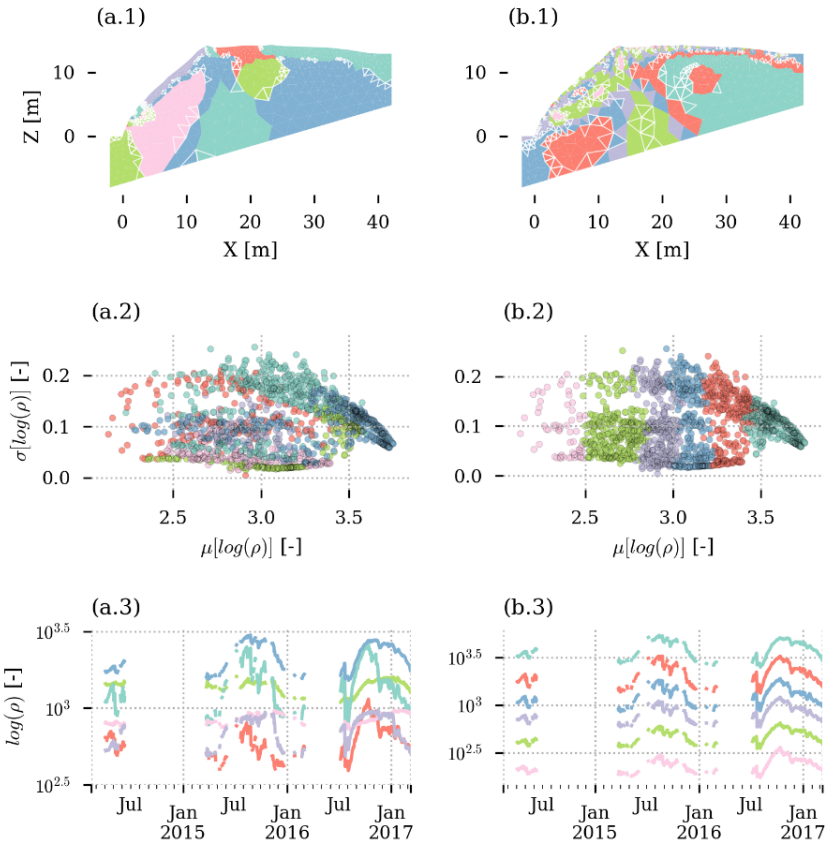

Figure 5-4: Diagnostic plot for HAC clustering (k=6). (a) On the z-standardized inverted log-resistivity data. (b) On the inverted log-resistivity data. The first row shows the spatial representation of the clustering partition with the cells having negative $S(i)$ values (Eq. 5.3) displayed with white edges; the second one their distribution in the scatterplot of the mean log-resistivity $\mu[log(\rho)]$ versus its standard deviation $\sigma[log(\rho)]$ for each of the 1558 ERT series ; the third one shows the averaged log-resistivity time-series per clusters.

Ideally, an adequate clustering method for the recovery of hydrofacies should leverage both information about the correlated dynamics from the z-standardized data and the raw resistivity. For instance, this could be done either by considering the raw log-resistivity in the clustering processes with some weighting scheme or by defining a posteriori a consensus clustering (see Monti et al., 2003) between a.1 and b.1 in Figure 5-4. Such methods were not developed because of the inability to validate or identify the number of clusters with such a dual approach that involves a wide range of potential compromises. This opportunity would rather be investigated using virtual experiments, which falls beyond the scope of this study. Notwithstanding, Figure 5-4 portrays the clustering results in a way that allows a fine diagnostic of the outcome and, eventually, a supervised reclassification of the groups when intra-cluster resistivity ranges are too broad.

5.3.2 Spatially Constrained Clustering

To mitigate the inconsistencies brought by the wide ranges of mean log-resistivity and standard deviation (Figure 5-4.a.2) within clusters, a first possibility is to disjoint those that are spatially (Figure 5-4.a.1) or statistically (Figure 5-4.a.2) split. Another possibility is to spatially constrain the clustering by providing a spatial connectivity matrix to the HAC algorithm (see section 5.2.2.2). The constraint will increase the number of the cluster over six, up to the point that the partition is both spatially and dynamically consistent in terms of correlation. The process of selecting the appropriate number of clusters k with the Silhouette Index (SI, Eq. 5.4) is repeated in Figure 5-5.a using HAC on the z-standardized log-resistivity data. Respectively, the blue and the orange curves report the SI with and without the use of the spatial connectivity constraint. The AMI similarity (Eq. 5.7) between the two approaches is given by the green curve. Above k=6, a first local optimum appears at 9 clusters. With k=9, the spatial organization of patterns (Figure 5-5.c.) is comparable to what is seen in Figure 5-3 or Figure 5-4, but the top slope, here in violet (zone C in Figure 3-5), has merged with a wider fractured area (Figure 3-5, zone H). The latter is poorly defined as being fully characterized by negative SI values (Figure 5-5.c. white edges). This means that the dynamics found in this area are more similar to those of other clusters, mainly the turquoise one in Figure 5-5.c if compared to Figure 5-3 or Figure 5-4.

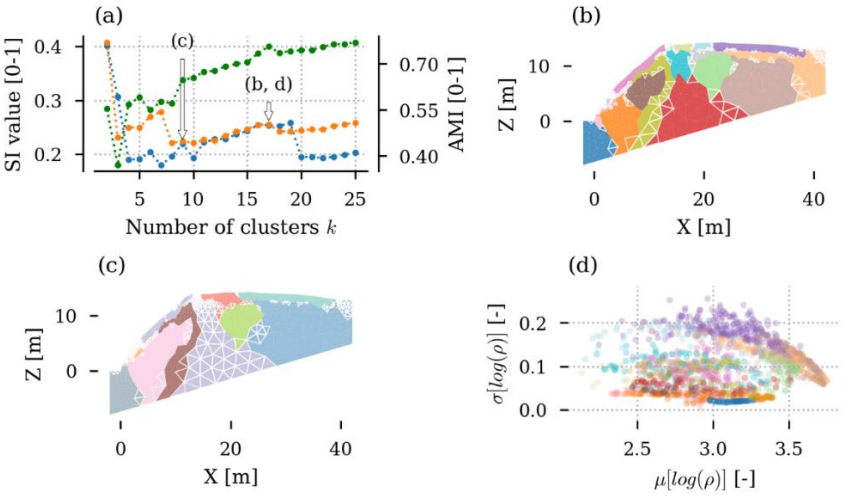

Figure 5-5: Selection of the number of clusters for the HAC method with connectivity constraint. (a) Silhouette Index (*SI*, Eq. 5.4) for the HAC with connectivity constraint (blue), without it (orange), and their similarity (green) given by the Adjusted Mutual Information (*AMI*, Eq. 5.7). (b) HAC with connectivity constraint and k=17 clusters. (c) HAC with connectivity constraint and k=9. (d) Scatterplot of the mean log-resistivity $\mu[log(\rho)]$ versus its standard deviation $\sigma[log(\rho)]$ for the clustering presented in (b). Cells with negative Silhouette values in (b) and (c) are showed with white edges.

Further apart, a better optimum is found around 17 clusters (Figure 5-5.a), which coincides with a small peak in the *AMI*, and equivalent *SI* in both clustering approaches. Thus, we have interpreted this point as a methodologically consistent number of clusters. Above 19 clusters, the *SI* with connectivity constraint drops to 0.2. With k=17 (Figure 5-5.b), the main former spatial patterns remain recognizable with the notable difference that a second horizon appears in the plateau, as in Figure 3-5 (zone B) or Figure 5-3.i. Another difference is that the pink cluster of Figure 5-5.c has been split into two parts. Besides, the spatial constraint has the effect of restoring more consistent groups in terms of average resistivity and standard deviation (Figure 5-5.d). A part of this consistency is, however, explained by spatial smoothness constraint in the inversion scheme. Finally, k=17 results in the presence of many small clusters, mainly in uncertain areas (see Figure 5-3.h and Figure 5-4.a.1) at the bottom of the slope or above the fractured area. However, if the smaller groups and those located relatively far from the surface electrodes are ignored, the partition would provide about ten spatially distinct groups that are geophysically interpretable, similarly to Figure 3-5.

Still, it appears that the more conductive porous limestone area (Figure 3-5, zone E) does not appear even with many clusters as high as 17.

5.3.3 Sensitivity and Robustness of Clustering Partitions

Figure 5-6 shows clustering applied to various time-series representations that are or not log-scaled, normalized, differenced, or decomposed (section 5.2.1). The first block (a to h) applies HAC with k set to 6 clusters while the second one (i to p) considers 9 clusters with a spatial connectivity constraint. Within each block, the two rows represent the choice to work either on the resistivity (Ω.m) or its log-transformation. Then, the clustering is applied, respectively to the columns of Figure 5-6, on these raw datasets (Raw data), the z-standardized (Eq. 5.1) ones (Z-std data), their first order differences followed by a z-standardization (Diff & Z-std data), and finally on the five first components of the PCA decomposition of the z-standardized data. Each labeled pair of figures represent the Silhouette index (SI, Eq. 5.4) as a function of the number of clusters k, and below it, the spatial patterns on the bottom related to k=6 or 9, whether a spatial connectivity constraint is considered or not, in phase with Figure 5-4.a and Figure 5-5.c. Although the selected k is not always a local optimum, which varies across time-series representation, arbitrarily fixing the number of clusters allows comparing the similarity of partitions with the Adjusted Mutual Information (AMI, Eq. 5.7). The references are the clustering applied on the z-standardized log-resistivity (b and j), which therefore have an AMI of 1. An AMI of zero reflects the score of two random partitions. In Figure 5-6, an AMI reaching 0.7 shows comparable spatial patterns with the reference. Of course, visual differences in the top of the model and closer to the electrodes have much more impact on the AMI score due to the variable resolution of the grid.

The new representation based on differencing (Diff & Z-std) produces interesting cluster distributions in terms of spatial patterns. Without spatial constraint, the clusters are nevertheless mostly continuous. The massive limestone area on the right of the fractured area (Figure 3-5, zone F) does not appear. The differencing removes the seasonal variation of the resistivity (see Figure 5-4.a.3), making this spot more synchronous with the rest of the limestone area below the plateau. The zone is, however, identified when the connectivity constraint is applied. Another interesting cluster is the banana-shaped one below the slope surface corresponding to the area of clayey

limestone (G in Figure 3-5). The shape maps well with the ones retrieved from raw resistivity data (a, e, i, m), indicating a consistent cluster.

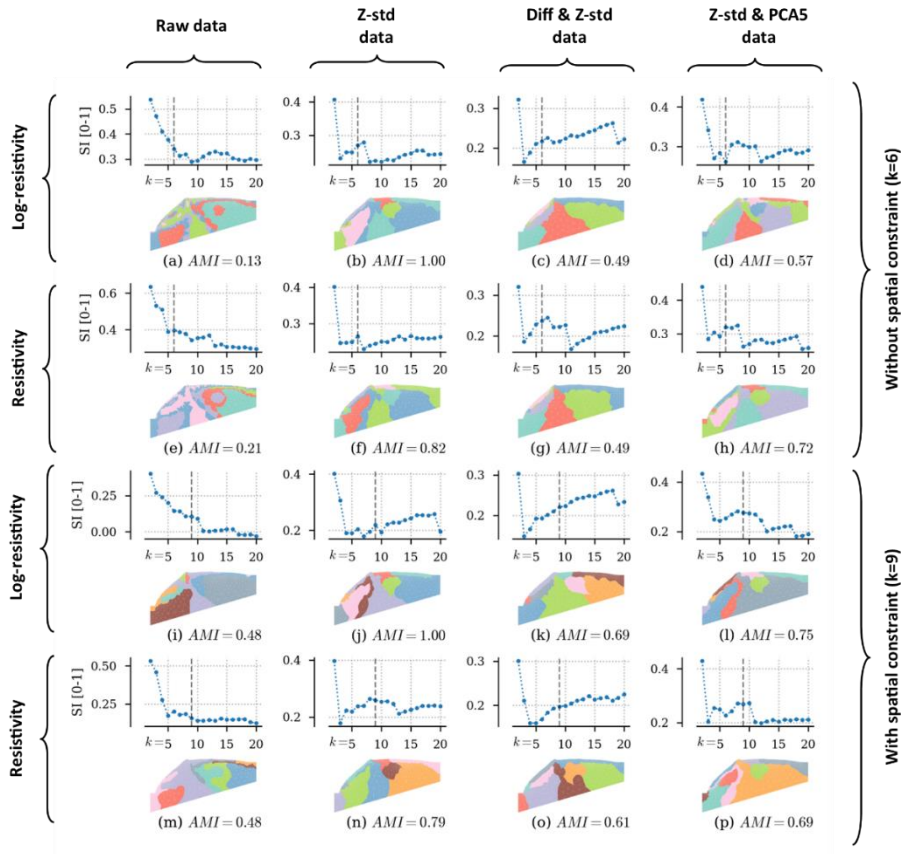

Figure 5-6: HAC clustering applied to various time-series representation. (a to h) With k = 6 and (i to p) with k = 9 clusters and a spatial connectivity constraint. Time-series representations considered for the clustering are either the resistivity or the log-resistivity as raw data (column one), z-standardized data (column two), differenced, and z-standardized data (column 3), and decomposed z-standardized data into 5 PCA components. Each label (a to p) shows the Silhouette variation according to the number of cluster k and the spatial patterns of clusters for the k indicated by the vertical dashed line in the Silhouette plot.

Regarding the application of the PCA (Z-std & PCA5), the clustering produces similar clusters compared to the reference, except for (d), but the value of 6 clusters does not seem appropriate given its sub-optimal SI. Otherwise, not much information is lost from the decomposition, and this option could be considered for reducing the computational requirement of the clustering task.

Yet, all clustering tasks shown in Figure 5-6 are applied to the full time-span of the dataset, 465 days. Another aspect of sensitivity is related to the question: how much information (i.e., days) is necessary to retrieve the clustering partitions of Figure 5-6? The question is addressed in Figure 5-7 with HAC clustering applied with and without connectivity constraint on the log-resistivity data and its four representations shown in Figure 5-6. The selected days are sampled according to two strategies. The first one (a and c) picks random but different (without replacement) ERT samples meaning that the samples could be spread over the full time-span of 465 days. The other strategy (b and d) picks continuous samples (i.e., consecutive days). For each given size, the sampling is repeated 50 times. The *AMI* is computed between each of the 50 clustering outputs and the partition obtained with the same time-series representation on the full time-span of 465 days (Figure 5-6, a to d and i to l).

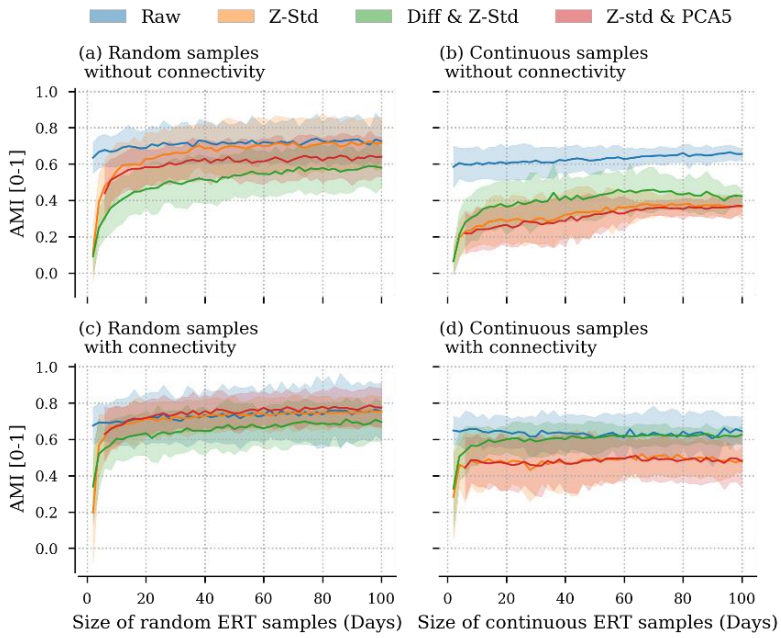

Figure 5-7: Convergence of HAC clustering partitions for various representations of log-resistivity data with increasing size of the sample sets. *AMI* (Eq. 5.7) is computed between every 50 runs of clustering on the sample sets and the partition retrieved on the full dataset of 465 days: (a) with random samples (k=6); (b) with continuous samples (k=6); (c) with random samples and with connectivity constraint (k=9); (d) with continuous samples and with connectivity constraint (k=9). The curve presents the mean and the 2 standard deviation bands for each representation of Figure 8 (Raw, Z-std, Diff & Z-std, Z-std & PCA5).

According to the random sampling strategy (Figure 5-7, a and c), the clustering applied on samples of the raw log-resistivity (blue) provides stable AMI across the range of sampling sizes (2 to 100 days). Compared to the full dataset, similar clustering partitions with high $AMI \cong 0.7$ are obtained even on small sample sets, with or without connectivity constraints. It means that there is not much added-value of a long time-span when the clustering of raw resistivity is performed. Therefore, it does not matter much if the sample sets are continuous or not. Regarding the clustering on the z-standardized data (orange) and decomposed data (red), the convergence of the AMI mostly occurs with samples sets below 20 days with k=6 and without connectivity constraints. The decomposed data (red) has a lower convergence limit (~ 0.6) compared to the z-standardized data (orange). This is because of the unstable 465 days patterns retrieved with PCA (Figure 5-6, d), which was the reference for computing the AMI. Convergence is faster and occurs mostly between 10 days when the clustering is applied with k=9 and a connectivity constraint (Figure 5-7.c). However, there is a drop in the AMI limit when the clustering is applied on continuous sample sets (Figure 5-7, b and d), and AMI does not exceed 0.5 even with a time-span as significant as three months. Since it is not the case with random sampling (Figure 5-7, a and c), one may conclude that the clustering applied on the full dataset is mostly based on seasonal variation, as shown in Figure 5-4.a.3. This behavior is different for the differenced dataset (green) as the seasonal variation is removed by the differencing. Consequently, with random sampling (Figure 5-7, a and c), it converges less rapidly and to a lower AMI compared to the z-standardized (orange) and the decomposed (red) dataset. The loss of AMI is also lower when it comes to continuous sampling (Figure 5-7, b and d).

Regarding the continuous sampling strategy, the drop of AMI for the z-standardized dataset (orange) and the decomposed ones (red) may raise several concerns related to the geophysical investigation and the methodology for recovering hydrofacies from correlated dynamics. Indeed, Figure 5-8 shows the different clustering partitions obtained from four different continuous periods of 20 days. The spatial patterns substantially deviate from the usual one retrieved on the full dataset. On the one hand, it may suggest that the patterns retrieved on the z-standardized are not robust unless applied on a long term ERT dataset covering at least a year, given the daily measurement strategy and the seasonal patterns shown in the data (Figure 5-4.a.3). On the other hand, the results simply reflect changes through time in

dynamically correlated features across space, which may include changes in the optimal number of clusters.

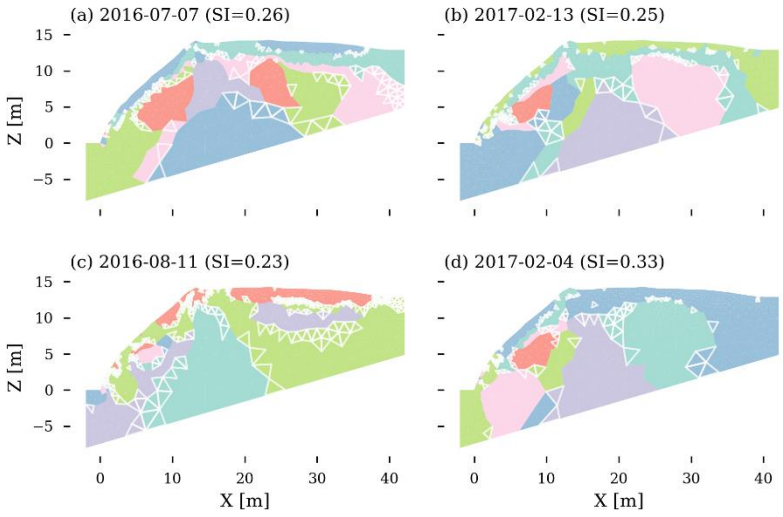

Figure 5-8: HAC clustering (k=6) applied to four continuous samples of 20 days of log-resistivity. (a to d) shows the results from four different starting days with the Silhouette index (SI, Eq. 5.4) reported in parenthesis. The negative Silhouette values ($S(i)$, Eq. 5.3) are displayed in with white edges.

From that point of view, hydrofacies may change over time according to the hydrological states of the systems, water distribution, and the patterns of hydrological connectivity. The result of Figure 5-8 may portray some of these changes. However, this chapter's focus is generally speaking on the robustness of clustering methods, and no attempt was made to engage in premature hydrological interpretation. Instead, the point is to underline the sensitivity of the method and the need to develop more robust methods for the clustering of hydrofacies, for instance, by considering the raw resistivity or other geophysical data in the process. Finally, it is worth recalling that the difficulty of the clustering task in this particular case is linked to the karstic site's complexity. The retrieval of coherent groups may be more comfortable in a less heterogeneous environment.

5.4 Conclusion

Nowadays, computer-assisted vision is increasingly used to extract and delineate geological and hydrological features, sometimes referred to as litho or hydrofacies, from ERT models. While early studies provided applications for non-time-lapse ERT models, applications to time-lapse models are still underrepresented. On short time-lapse models (< 20 days), Genelle et al. (2012) and Xu et al. (2017) developed the first applications based on time-series clustering (TSC), assuming that these structures can be extracted based on the similarities observed in the time dynamics of resistivity. The basic clustering principles were introduced, together with three clustering evaluation metrics and one clustering similarity metric. Using a 465 days time-lapse ERT model of 1558 cells acquired from the surface of a heterogeneous karstic environment (Watlet et al., 2018b, 2018a), this chapter studies: (1) the comparison between the three clustering algorithms k-means, hierarchical agglomerative clustering (HAC), and Gaussian Mixture Model (GMM), including the question of the optimal choice of cluster number and the identification of potentially misclassified spatial cells, (2) the effect of adding a spatial constraint in clustering, and (3) the robustness of the clustering outputs to various representations of the resistivity data as well as the impact of the number of days considered in the ERT model for the clustering task.

Specifically, applied to 1558 z-standardized resistivity series of 465 days, the three candidate algorithms produce similar spatial patterns that highlight temporarily correlated areas across space. Six clusters were considered based on our clustering evaluation metrics. However, such clusters may be spatially split and may include cells with substantial differences in their mean raw resistivity or standard deviation. Hence, clustering based on the correlation of resistivity series obtained from z-standardized data may retrieve geologically inconsistent groups. Alternatively, clustering on the raw resistivity time-series is dominated by their mean resistivity. Accordingly, the retrieved clusters depict iso-resistivity areas, but their averaged temporal dynamics are all correlated, and nothing is learned about the specific dynamic property of the subsurface elements. This, therefore, encourages to work on the standardized resistivity while checking the raw resistivity distribution within clusters.

In the second part, the HAC specificity of adding a spatial constraint was considered such that clusters are spatially tied into one feature. The constraint had the expected effect of increasing the suggested number of clusters to 9 or

17. With 9, one of the clusters would have needed to be separated for consistency. With 17, the expected patterns are well represented by about ten clusters. The remaining clusters were either relatively small or distant from the electrodes, thus deserving less consideration. The results were more consistent in their raw resistivity and standard deviation while applied on the z-standardized data due to the spatial proximity of the cells, but some raw resistivity patterns are not always revealed from correlation patterns.

In the last section, HAC with and without connectivity constraint was applied to 8 different time-series representations where the resistivity is, or not, logarithmically scaled, standardized, differenced, or dimensionally reduced with principal component analysis. The major differences in spatial patterns remained between the raw resistivity and the other representations revealing correlated areas. The redundancy of patterns across the different representations creates confidence in the patterns that are restituted. However, the sensitivity analysis based on smaller sample sets showed that these patterns are associated with the seasonal dynamics of resistivity and cannot be retrieved from the standardized data even with continuous sample sets of 100 days. It also shows how much interpretation can vary between a single ERT survey and time-lapse experiments and from one short-term time-lapse survey to another. Still, less than 20 days are necessary to retrieve the long-term patterns if they are not continuous but randomly picked in the model. This last result may encourage long-term ERT monitoring of at least one year to retrieve robust clusters. It may also depict the temporal variability of water distribution, hydrological processes, and so hydrofacies if they are identified from short-term correlated resistivity.

In general, the results encourage to perform clustering of time-lapse ERT models with various numbers of clusters, various time-series representations, and various sample sets to gain confidence from redundancies between the resulting patterns. Redundancies, as cluster evaluation indices, support decision making. Shortly, more robust clustering methods for the identification and zonation of hydrofacies and lithofacies will benefit from integrating both the information about raw resistivity and temporal dynamic similarity, and eventually other geophysical datasets (e.g., Di Giuseppe et al., 2014, 2018; Paasche et al., 2006). Regarding the algorithms, HAC was particularly interesting for its versatility. Besides its ability to constraint spatially the clustering, HAC can be applied with any distance metrics. If HAC does not account directly for the uncertainty in the clustering such as GMM,

such uncertainty may be computed from bootstrap samples in the particular case of time-lapse ERT datasets. In that spirit, HAC is used to generate consensus clustering that may already be helpful to create a final clustering from several clustering partitions (Monti et al., 2003). Further guidelines will most likely be fruitfully developed in combination with synthetic experiments combining resistivity and hydrological modeling since the main difficulty of clustering is its unsupervised nature and the difficulty of appreciating the validity of the outcomes.

Chapter 6 Inferring karst hydrology from time-series using causal inference methods

"That a few Original Ideas may be made to signify a great number of Effects and Actions, it is necessary they be variously combined together: And to the end their use be permanent and universal, these Combinations must be made by Rule, and with wise Contrivance."

Berkeley

Foreword

This chapter covers the central theme addressed by the MIGRADAKH project (section 1.2.1), namely causal inference from time-series. As previously reported, the illogicalities of the initially chosen method, Convergent Cross Mapping (CCM, Sugihara et al. 2012), led me to believe that the initial assumptions (or beliefs) in the thesis proposal were ill-founded. From this disillusion, I sought to understand complexity (section 1.1) and wanted to broaden the framework and definition of causality (Chapter 2). Without abandoning time-series analysis, I explored causality from the perspective of identifying patterns of complexity through nonlinear time-series analysis (Chapter 4) and spatial patterns of dynamic similarity (Chapter 5).

Meanwhile, I have refined my expertise and understanding of causal inference methods. In March 2018, I visited Prof. Ray Huffaker from the University of Florida that has developed a framework for causal inference based on the extraction of deterministic components of time-series using relying on CCM (Huffaker et al., 2017). In June 2019, we developed a Python version of the CCM algorithm with Dr. Olivier de Viron at the University of La Rochelle. I also turned to other methods and applied them with other students and researchers (Brulein, 2019; Delforge et al., 2019; Got, 2019; de Viron et al., 2019). The analysis presented at the AGU fall metings 2019 (de Viron et al., 2019) is currently extended in a paper entitled *Causal relationships in the climate system* submitted to the Journal of Climate. Its content, however, does not fall in the scope of this thesis. This chapter is a much more detailed version and expansion of the analyses presented at the general assembly of the

European Geoscience Union 2019 (Delforge et al., 2019). This content is undoubtedly one of the most detailed introductions to applied causal inference in hydrology, showing these methods' potential, the variability between methods, and the critical necessity to think about their hypotheses, parameters, and input data. Selected materials focused on hydrological connectivity from this chapter will be wrapped into a paper to be submitted in a scientific journal.

Regarding causality, causal inference methods relate to the formal cause (d8, Appendix III) by portraying the account of what is to be in terms of constant conjunctions (d24, d30) and reporting a system's organization in a causal graph, model, or frame (d34). However, we expect these methods to bring to light interactions and dependencies that result from existing processes, i.e., cause-effect relationships (d1, d9). The most commonly used method for causal inference remains the linear correlation. With time-series analysis, the cross-correlation function relies on the principle of priority (d22). Similarly, CCM is also a bivariate method, but it considers nonlinear dependencies between time-series (d8, d24). CCM evaluates them using a nearest-neighbor regressor such as EDM-Simplex (Chapter 4). Beyond dependencies and their predictive skills, CCM relies on the principle of convergence to assess causality. The convergence criterion is an original and optimistic instrumentalization (d27, d32) of the *problem of induction* (section 2.4.3) to infer causality. It tells us that our understanding of causality, if causality there is, should increase with more data. It is also a kind of asymptotic teleology (d39). Causality could then be measured through an observed increase in the prediction skills (or in robustness, such as in Chapter 5, Figure 5-7). However, if convergence is necessary and potentially sufficient to claim that two variables belong to the same dynamical system, it is not sufficient to claim a direct cause-effect relationship or a hydrological connection. Indeed, CCM does not consider Reichenbach's principle and common drivers (d26, Appendix III). That is why this chapter also investigates methods that do, using conditional independence as a criterion for causality (Runge et al., 2019a). The chapter shows that the most sophisticated method is the most robust in virtual experiments (d17). However, on real and incomplete datasets (d16), it lacks robustness and shows hardly interpretable results, in which case relying on simpler – and, therefore, potentially inadequate - methods could bring more inference power following the logic of parsimony (d18).

Abstract

The chapter focuses on detecting causal relations in complex hydrogeophyiscal systems, in this case, karstic systems. The objective is to compare four causal detection methods: two bivariate methods, the linear cross-correlation function (CCF) and the nonlinear convergent cross-mapping (CCM), and two multivariate frameworks based on conditional independence. The first one relies on partial correlations (ParCorr), the second one on conditional mutual information (CMI). The methods are applied using three levels of analysis: (1) a virtual experiment involving two disconnected hydrological reservoirs, hence not causally related, but both forced by effective precipitation data; (2) a case study focused on the detection of connected preferential flows between the surface and cave percolation in the Rochefort cave system; (3) a case study focused on the general functioning of the system including relative gravimetry data.

The virtual experiment (1) recommends using bivariate methods of differenced data to better spot time-dependencies between hydrological variables. Multivariate methods should be applied to the raw (not differenced) data. Compared to ParCorr, the same experiment suggests that CMI is the most robust method to identify effective hydrological connections, as it detects no spurious link between the two reservoirs. CMI was, however, not robust on the real hydrological datasets (2 and 3), possibly due to dynamic artifacts in the Electrical Resistivity Tomography (ERT) model of the subsurface (in case 2) and relatively short time-series overlap due to missing data (in both cases). ParCorr and CCM showed more intuitive causal structures and a higher agreement level in the causal links. They allowed spotting fewer causal associations than the conventional CCF approach that detects ubiquitous causal associations due to the global forcing of meteorological variables. Encouragingly, they spotted strong associations between surface resistivity patterns and cave drip discharge (2), where the effective connectivity was confirmed by dye tracing. Whatever the method and the experiment, spurious links are often recovered, and the possibility of constraining the direction of causal links was investigated. The first real case (2) showed that constraining had not much effect on ParCorr results beyond the constraint, while CMI remained unstable. In case 3, the constraint allows retrieving systematic time-dependencies of atmospheric pressure effects on relative gravimetry both with ParCorr and CMI.

In conclusion, other causal inference methods beyond CCF are recommended as they outcome a more intelligible causal structure by focusing on fewer causal links. Although virtually and theoretically the most robust, CMI may not be suitable for short hydrological time series of variable quality or data generated or corrected by a model. Importantly, p-values do not relate to the probability of causal associations. The results vary substantially with different hypotheses materialized in datasets, methods, parameters, or user constraints. Accordingly, multiple approaches are encouraged for a better inference and confidence in the results when some links are found redundant. Any approaches should be transparent and justified in their context because, while empirical, they remain based on methodological and user assumptions.

6.1 Introduction

Modeling in hydrology is predominantly based on a hypothetico-deductive or physically-based approach (section 2.3.3.2, and 2.4.2). The hydrology of a system is derived by combining structural data (e.g., topography, pedology, geology), meteorological data (e.g., rainfall, evapotranspiration), mechanisms representing hydrological processes, and, if available, calibration data. An ideal model is supposed to reveal water distribution, flow paths, and velocities within the system. While this approach could also be applied to karst systems (Hartmann et al., 2014; see also Figure 1-3), physically-based models impose strong and unverified assumptions on karst systems that are complex to understand (section 1.1.2, 1.2.2.1). Bakalowicz (2005) reminds the limits of relying upon structural data to characterize their hydrological functioning. In many cases, the structure and heterogeneity of a karst system remain hidden. They are not sufficiently characterized to deduce processes, including fast preferential flows resulting in the high hydrological responsivity of karst systems.

As a result, karst investigation often relies on a panel of functional and empirical approaches (see Bakalowicz, 2005), in particular, time-series analysis. Cross-correlation analyses are the most common, frequently combined with spectral or wavelet analysis (Angelini, 1997; Bailly-Comte et al., 2008; Kadić et al., 2018; Labat et al., 2000; Larocque et al., 1998; Mathevet et al., 2004; Mayaud et al., 2014; Ollivier et al., 2015; Padilla and Pulido-Bosch, 1995; Schuler et al., 2020; Tagne and Dowling, 2018). Bivariate cross-correlation analyses are particularly suited to identify and detect the occurrence of rapid preferential flows in karst systems. Peak-to-peak times could be interpreted as causal delays. However, peaks in correlograms may reflect superimposed processes. The system's general lumped behavior, such as the rainfall-spring discharge relationship, may not represent the actual flow path and hydrological connections within the system.

Retrieving connectivity from bivariate time-series analysis is not trivial given that meteorological forcing acts as a confounding exogenous factor making endogenous hydrological variables all temporally related. Hence, as for correlation, cross-correlation does not imply causation. Linear time-dependencies assume linear processes and are consequently not sufficient to

infer causation. They also do not remove the effect of the common causes (section 2.4.3.2, d26 in Appendix III).

Modern causal inference methods now include either or both nonlinear approaches and multivariate frameworks that account for common causes to infer causation from time-series (e.g., Granger, 1969; Hlaváčková-Schindler et al., 2007; Runge et al., 2019a; Schreiber, 2000; Sugihara et al., 2012). They have evolved from the linear cross-correlation function to nonlinear and multivariate frameworks for causal discovery and are gaining popularity in Earth Sciences (e.g., Goodwell et al., 2020; Meyfroidt, 2016; Runge et al., 2019b; see also Figure 2-1; Appendix I).

Recently, hydrological studies have focused on these methods to infer hydrological connectivity (Rinderer et al., 2018; Sendrowski and Passalacqua, 2017; and section 2.4.3.2). In particular, Rinderer et al. (2018) distinguish three types of connectivity: (1) the static and structural connectivity, which highlights fully potential paths for water flows from geomorphology; (2) the functional one, which could be derived from hydrological data and their bivariate time-dependencies, (3) and the effective connectivity, which is supposed to reveal actual flow path within the system, possibly from tracing test or more advanced causal inference methods.

In this chapter, four causal inference methods, as a combination of linear/nonlinear and bivariate/multivariate methods, are selected and compared. Three different cases allow discussing the potential of causal inference method to unravel effective connectivity and distinguish it from the functional patterns of connectivity. The first case study illustrates and evaluates these methods' potential by considering a model with two disconnected reservoirs, responding in parallel to effective precipitation. Hence, the two reservoirs present functional patterns of connectivity but are not effectively connected. The methods are applied in order to check if they are able to report this absence of effective connection. The next two cases apply to the Rochefort cave datasets. Focused on vadose zone preferential flows, the first one aims to relate subsurface resistivity patterns (section 3.3.2; Chapter 5) above the cave to drip discharge dynamics within the cave, including rainfall and potential evapotranspiration in the analysis (section 3.3.3; Figure 3-6). Drip discharge data are monitored locally in three locations within the cave (section 3.2; Figure 3-2). Previous dye tracing tests have revealed fast preferential flow between the surface a particular spot in the cave

(P1, Figure 3-2), making it a good study case to assess if causal inference methods may detect the effective connection. The last case is a complementary study focused on how causal inference methods depict the system's general and functional behavior. In particular, the causes of the mass balance in the cave system (monitored by relative gravimetry data) are inferred and discussed considering other variables: two drip discharge time-series, and the potential drivers of gravity: atmospheric pressure, rainfall, evapotranspiration, and groundwater level (Figure 3-6).

6.2 Theory and Methods

6.2.1 Causal Inference Methods

The selected methods apply to the time-domain (see Runge et al., 2019 for a broader review) and investigate causal relationships between time-series over a time window defined by a maximum time delay d_{max}. Therefore, they do not investigate causal relationships beyond this delay and may be inappropriate when the processes under study substantially alter the frequency of the dynamics between the driving variable and the response variable. Most hydrological processes act as a low-pass filter. The water's multiple paths and their difference in flow velocity characterize the dispersivity of porous and heterogeneous environments. The dispersivity is supposed to be lower in a karstic environment if one considers only the highly responsive preferential flows within fractures, conduits, or macropores (e.g., Hartmann et al., 2014). The causal inference methods will be applied to discover such fast flow occurring within a few days in a karstic vadose zone and causal relationships in the general karst system monitored by gravimetry.

Four methods of causal inference are used: (1) the Cross-Correlation Function (CCF), (2) the Convergent Cross-Mapping (CCM) method, (3) the PCMCI-Partial Correlation method (ParCorr), and (4) the PCMCI-Conditional Mutual Information method (CMI). They are all based on the principle of priority of the cause (d22 in Appendix III). It means that the cause must occur before its consequence. Formally, addressing causality is identifying a robust association between a response variable Y_t and a time delay d of its potential driver X_{t-d}. Delays provide a time-asymmetry such that causal interactions can be reported in the form of a graphical model or Directed Acyclic Graph (DAG) (Pearl, 2000; Spirtes et al., 1993) with arrows linking driving variables to their responses variables (see Figure 6-1, similarly to Figure 1-1).

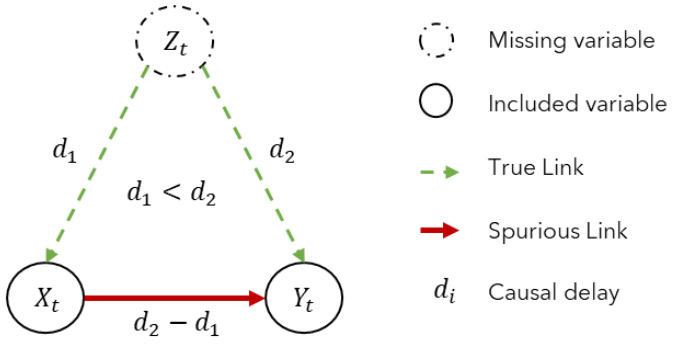

Figure 6-1: Example of incorrect DAG showing a spurious link $X_t \rightarrow Y_t$ due to a missing common driving variable Z_t. X_t is seemingly causing Y_t because X_t systematically responds before Y_t to Z_t.

The methods vary according to two criteria: (1) whether they are suitable for detecting linear or nonlinear causal dependencies, and (2) whether they provide a bivariate or multivariate analytical framework (Table 6-1).

Table 6-1: Classification of selected causal inference frameworks.

	Bivariate	**Multivariate**
Linear	Cross-Correlation Function (CCF)	PCMCI - Partial Correlation (ParCorr)
Nonlinear	Convergent Cross Mapping (CCM)	PCMCI – Conditional Mutual Information (CMI)

When faced with hydrological data, the question arises as to whether a linear or nonlinear method should be chosen. On the one hand, the hydrological response to effective precipitation is expected to be sensitive to the system's initial conditions (e.g., its water content and distribution), which suggest considering nonlinear methods. On the other hand, one expects that nonlinear relationships remain mostly monotonous, allowing linear thinking, e.g., the more rain, the more percolation. Linear methods are in phase with this linear thinking. They have the advantage of telling what change is expected in one quantity while changing another quantity, which is very interpretable. Nonlinear methods do not tell anything about the nature of the relationship, which could be, in fact, linear (see 1.1.2.3). For these reasons, both types of methods are considered.

A multivariate framework allows the common cause principle (d26 in Appendix III) to be taken into account, thus eliminating spurious causal interactions arising from confounding. Indeed, with the sole basis of the principle of priority, causal inference methods may identify spurious associations, such that $X_{t-d} \rightarrow Y_t$, when both variables are not directly causally related but share a common driver Z_t. This pathological case is illustrated in Figure 6-1.

Generally speaking, not considering an explanatory variable in the causal analysis may result in an incorrect DAG. Consequently, causal inference methods rely on the hypothesis of causal sufficiency, i.e., measured variables include all the common causes (Reichenbach, 1956; Runge, 2018a). Paradoxically, causal sufficiency favors the quest for data with the best spatial and temporal coverage and resolution. However, causal inference methods remain impractical on such data sets because they are subject to the curse of dimensionality and limited computational resources (Runge et al., 2019a). Causal sufficiency must be approached rather parsimoniously and include a dataset that is representative of the emerging and dominant patterns in the system, which can be obtained from a high-resolution dataset using dimension reduction approaches (e.g., Chapter 5).

6.2.1.1 Cross-Correlation Function (CCF)

The cross-correlation function (CCF) is the most common method to analyze linear time-dependencies and address causality, including for karst systems (e.g., Bailly-Comte et al., 2008; Mathevet et al., 2004; Ollivier et al., 2015; Schuler et al., 2020; Tagne and Dowling, 2018; Watlet et al., 2018b). For a driving variable X_t and a response variable Y_t, causality is inferred by computing the cross-correlation function (CCF) and from the principle of priority of the cause (d22 in Appendix III). For a window of absolute delays $[0, d_{max}]$ with $d_{max} \geq 0$, Pearson's correlation coefficient ρ is computed between the response and the delayed driver on their overlapping domain:

$$CCF(d) = \rho(X_{t-d}, Y_t) \qquad Eq.6.1$$

with $d \in [0, d_{max}]$.

Pearson's ρ between two time-series X_t and Y_t is the ratio between their covariance and the product of their standard deviations:

$$\rho(X_t, Y_t) = \frac{cov(X_t, Y_t)}{\sigma_X \sigma_Y} \qquad Eq.6.2$$

The ρ coefficient is a standardized measure of linear dependencies that can be interpreted as the slope of a linear regression between the two z-standardized variables (Eq. 5.1). Accordingly, ρ is ranging between -1 and 1, meaning respectively perfectly anti-correlated or correlated. A ρ of zero indicates the absence of linear dependencies. The significance of the hypothesis that ρ is different from zero is usually assessed analytically through a Student's-t test reporting a p-value. The p-value estimates the probability that ρ is the output of an uncorrelated process. The p-value is sensitive to the number of overlapping samples such that more samples are required to have a significant p-value if $|\rho|$ is low. For a significance level α, significant relationships have p-value lower than α. In this chapter, significant correlations and their causal delay d are reported in a DAG. The case of $d = 0$ does not allow to infer a direction for the causal relationship and are reported with bidirected arrows. In general, CCF is symmetric if computed between $[-d_{max}, d_{max}]$ but the sign of ρ allows to interpret the results and conclude if there is a transfer of water.

6.2.1.2 Convergent Cross Mapping (CCM)

Convergent Cross Mapping (CCM) is a causal inference method rooted in nonlinear dynamical systems theory. It allows detecting weak nonlinear associations between two time-series (Sugihara et al., 2012). CCM is part of the Empirical Dynamic Modeling (EDM) framework and is a bivariate extension of the EDM-Simplex algorithm (section 4.2.2). Hence, CCM similarly relies on Takens's embedding theorem (see Figure 1-4, Eq. 4.1). To address whether X_t causes Y_t, the response variable Y_t is first embedded using Takens's state space reconstruction. The optimal embedding parameters (m and τ, see also Table 4-3) for Y_t can be selected by optimizing the EDM-Simplex forecasting skills (Sugihara and May, 1990). From the optimal reconstructed state space M_Y of Y_t, CCM makes forecast of X_t based on reference time indices using practically the same workflow as presented in Figure 4-3. The difference in CCM is that the nearest-neighbor states for the times of reference found in M_Y, that are $\{\dot{y}_{d,1},\ \dot{y}_{d,2}, \dots, \dot{y}_{d,k}\}$, are mapped onto

X_t based on their time indices. Afterward, the forecast of X_t follows the same procedure. The difference is portrayed in Figure 6-2.

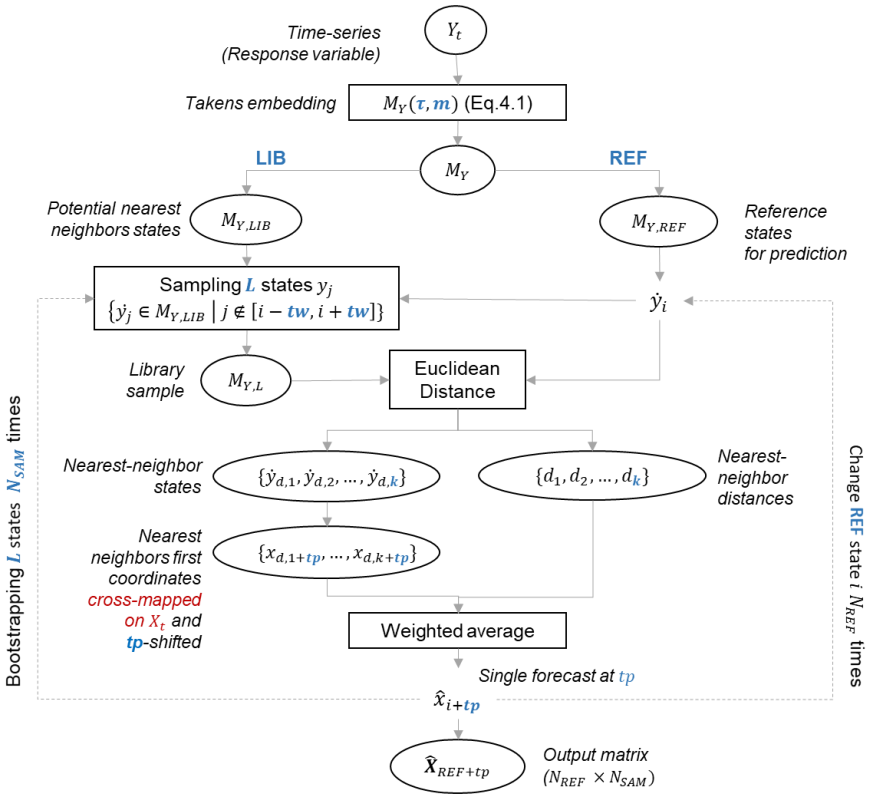

Figure 6-2: EDM-CCM algorithm flowchart. EDM-CCM is an extension of the EDM-Simplex algorithm allowing to make bivariate forecast and test for causality between a response variable Y_t and its driver X_t. The difference with the EDM-Simplex algorithm (Figure 4-3) is displayed in red. User-defined inputs are the same and described in Table 4-3.

The CCM philosophy may seem counter-intuitive. Indeed, the predictive arrow is opposite to the causal arrow being tested: if X_t causes Y_t, predictions are made from Y_t to X_t. The theory of nonlinear dynamical systems justifies this strategy. If X_t is driving Y_t, M_Y contains the information about the states of X_t (Figure 1-4). Nearest neighbors in M_Y should identify when the system's state is relatively similar to the reference state. Then, these nearest-neighbor time-indices obtained from Y_t (that are remote in time) are also a good basis for predicting points in X_t from the state of X_t at the same time-indices. This is true if and only if both variables belong to the same dynamical system. In other words, CCM goes beyond linear correlation by checking if two variables

behave consistently when the system is under the same state. This account of state dependency makes the method nonlinear, and the state of the system is assumed to be well represented by the pseudo-states found in the embedded reconstruction M_Y of the response variable Y_t.

In the original CCM paper (Sugihara et al., 2012), forecast skills are assessed with the mean Pearson's correlation coefficient ρ (Eq. 6.2) of the N_{SAM} forecast vector and a prediction horizon tp of 0, i.e., without delay. To identify a causal relationship $X_t \rightarrow Y_t$, Sugihara et al. (2012) relies on the principle of convergence, meaning that CCM-averaged forecast skills of X_t should increase with larger sample sizes L. With an increasing L, nearest-neighbor states on Y_t are supposed to be more and more relevant, which should increase the forecast skills of X_t. In practice, forecast skills do not increase indefinitely but up to a plateau. This is due to noise or to the general fact that closer nearest-neighbor does not improve the forecast substantially above a given L. For weak nonlinear causal associations, CCM yield an asymmetric pattern of convergence, such that convergence is only observed for the case where $X_t \rightarrow Y_t$, that is Y_t predicts X_t. The asymmetry of convergence is, therefore, an interesting pattern for distinguishing the cause from the effect.

However, in this chapter, convergence will not be considered as a sufficient criterion for several reasons. First, if the forecast skills ρ are high and significant for a given length L, it is fair to assume that they have converged or being in the process of converging. Secondly, Sugihara et al. (2012) explain that convergence can be observed in both ways (Y_t CCM-predict X_t and X_t CCM-predict Y_t) for a unidirectional causal relationship $X_t \rightarrow Y_t$ in case of strong statistical dependency between variables. Locally, hydrological variables exhibit such strong coupling due to the general forcing of precipitation and evapotranspiration. Finally, the maximum mean ρ obtained with the highest L may depend on the time-series noise properties, which may be substantially different in this case that combines data from various sources and sensors, including an ERT model. Hence, this variability makes the asymmetric predictive skills pattern at $tp = 0$ a potential misleading source to infer the direction of the arrow of causation.

Instead, to deal with strong coupling, Ye et al. (2015) suggested varying the prediction horizon tp to discriminate the driver from the response based on the principle of priority of the cause (d22 in Appendix III). Cross-map skills are supposed to increase while forecasting the past of X_t from Y_t if they are

causally related. This is illustrated in Figure 6-3 for a bivariate linear model such that $Y_t = 0.2Y_{t-1} + 0.2X_{t-2} + 0.2\varepsilon$ where X_t and ε are standard Gaussian noise. The black curves shows that the predictive skills of Y_t predicts X_{t+tp} are high for a negative tp. Hence, X_t CCM-causes Y_t. This is not the case when X_t predicts Y_t (gray curves). Predictive skills are higher for a positive delay tp. This time asymmetry is assumed to be a better indicator than the convergence to infer the direction of the arrow of causation.

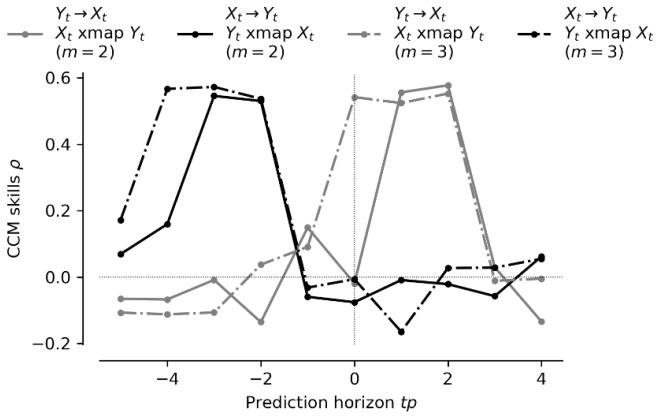

Figure 6-3: Example of CCM on a linear stochastic model $Y_t = 0.2Y_{t-1} + 0.2X_{t-2} + 0.2\varepsilon$ with a univariate causal relationship $X_{t-2} \to Y_t$. CCM skill ρ is the mean Pearson correlation between the $N_{SAM} = 100$ predicted vectors of length $L = 100$ and the corresponding true values. A time-delayed causal interaction is revealed if significant predictive skills are sustained at least over a window of m while predicting (xmap) the past ($tp \leq 0$) of the driving variable from the response variable.

Since the actual causal delay d is 2, Figure 6-3 also shows the effect of changing the embedding dimension $m = 2$ to $m = 3$ for an embedding delay τ of 1 (Eq. 4.1). The actual nonlinear (or linear) dependencies are sustained over $m - 1$ spurious delays. To avoid this effect, the continuous segments of significant dependencies should be truncated by $m - 1$ to draw a DAG from CCM dependencies. In doing so, only $tp = -2$ would be considered as a causal delay for the DAG of Figure 6-3.

The significance of the mean CCM skills ρ is also assessed through a Student's-t test. In the case where significant negative ρ values are reported, they are ignored since they have no meaning.

Regarding the other CCM parameters (Table 4-3), all analyses in this chapter rely on an embedding dimension of 2, a delay τ of 1 day. The number of nearest neighbors is the simplex $k = m + 1$, and the Theiler window tw is 10 days. All points that overlaps between the time-series are predicted (REF) 100 times (N_{SAM}).

6.2.1.3 PCMCI Algorithm: ParCorr and CMI

The PCMCI algorithm is described in Runge et al., 2019a, and implemented in the Tigramite Python package for causal time-series analysis (version 4.1 in this case). PCMCI is a 2-step procedure: PC, named after its authors (Peter Spirtes and Clark Glymour, 1991), and MCI standing for Momentary Conditional Independence.

In general, considering a multivariate time-series process $\boldsymbol{X_t} = \{X_t^1, \dots, X_t^p\}$ of p time-series, PCMCI allows recovering a DAG (e.g., Figure 6-1) based on conditional independence. Conditioning allows removing the effect of potential common causes (d26 in Appendix III). A delayed time-series X_{t-d}^i, with $i \in \{1, \dots, p\}$, is assumed not to cause itself or another one X_t^j, with $j \in \{1, \dots, p\}$, if they are independent conditionally to the past of the process $\boldsymbol{X_t^-}$, excluding X_{t-d}^i:

$$X_{t-d}^i \nrightarrow X_t^j \iff X_{t-d}^i \perp\!\!\!\perp X_t^j \mid \boldsymbol{X_t^-} \setminus \{X_{t-d}^i\} \qquad Eq.\,6.3$$

Hence, causal relationships are identified based on the rejection of conditional independence. For instance, the times-series X_t and Y_t of Figure 6-1 are not directly causally related given that $X_{t-(d_2-d_1)} \perp\!\!\!\perp Y_t \mid Z_t^-$, if X_t and Y_t have no self-dependencies.

The Full Conditional Independence algorithm (FullCI) is entirely based on Eq. 6.3. However, FullCI suffers from the curse of dimensionality if the conditioning involves too many variables in the set $\boldsymbol{X_t^-}$ (see Runge et al., 2019a). Hence, the purpose of the prior PC step is to estimate first the potential parents $\hat{\mathcal{P}}(X_t^j)$ for each variable X_t^j. Tigramite relies by default on the PC_1 iterative procedure. Initially, all potential parents are considered as $\hat{\mathcal{P}}(X_t^j) = \boldsymbol{X_t^-}$. In the first step, all the parents that are unconditionally independent with X_t^j are removed, and the parent presenting the strongest dependencies is

identified. In the second one, the parents that are independent to X_t^j conditionally to the strongest parent of step one are removed, and a second parent with the highest conditional dependence is identified as an additional condition for step three. The operation is repeated considering a 1-by-1 increasing number of conditions up to a point there are no more conditions to test in $\hat{\mathcal{P}}(X_t^j)$. Then, the MCI second step starts and tests for conditional independence on the dimensionally reduced sets of parents resulting from PC_1, such that:

$$X_{t-d}^i \nrightarrow X_t^j \iff X_{t-d}^i \perp\!\!\!\perp X_t^j \mid \hat{\mathcal{P}}(X_t^j) \setminus \{X_{t-d}^i\}, \hat{\mathcal{P}}(X_{t-d}^i) \qquad Eq.\,6.4$$

The test is conditioned both to the parent of the response $\hat{\mathcal{P}}(X_t^j) \setminus \{X_{t-d}^i\}$, and the time-shifted parents of the potential driver $\hat{\mathcal{P}}(X_{t-d}^i)$ to account for autocorrelation and to have a better estimate of the causal strength. Resulting from the MCI step, links where conditional independence cannot be rejected are considered as true causal parents, i.e., causally inferred sufficient causes. These links are reported in the DAG.

PCMCI flexibly allows us to consider different conditional independence tests: a linear method by assessing Partial Correlations (ParCorr) and a nonlinear one relying on Conditional Mutual Information (CMI). Both are used respectively as a linear and nonlinear multivariate framework for causal inference (Table 6-1).

Partial correlations are Pearson's correlations (Eq. 6.2) between X_{t-d}^i and the residuals of multivariate linear regression model of X_t^j against its conditions, for instance, $\hat{\mathcal{P}}(X_t^j) \setminus \{X_{t-d}^i\}, \hat{\mathcal{P}}(X_{t-d}^i)$. The linear model is fit using ordinary least square regression. The partial correlation significance is estimated with a Student's t-test (see 6.2.1.1), accounting for the increased degrees of freedom when necessary. This framework is very similar to the popular Granger causality (GC, see 2.4.3.2), which is based on vector autoregressive models (Granger, 1969). However, PCMCI-ParCorr differs from the usual GC in three aspects: (1) GC does not rely on the PCMCI procedure and, thus, suffers from the curse of dimensionality; (2) GC does not report contemporaneous dependencies ($d = 0$); and (3) GC relies on a F-test, testing if including a potential driver X_{t-d}^i in the multivariate model of X_t^j

significantly reduces the variance of residuals. In karst hydrology, very few applications involve partial correlations (e.g., Kadić et al., 2018).

In contrast, CMI can be seen as a multivariate extension of the transfer entropy method (Schreiber, 2000). In the information theory, CMI or $I_{X,Y|Z}$ as the mutual information between two variables X_t and Y_t conditioned to Z_t, is defined as:

$$I_{X,Y|Z} = \int\int\int p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}dxdydz \quad \text{, or} \qquad Eq.6.5$$
$$H_{XZ} + H_{YZ} - H_Z - H_{XYZ}$$

where H is the Shannon entropy (Shannon, 1948). If $I_{X,Y|Z} = 0$, X_t and Y_t are conditionally independent to Z_t, and, therefore, not directly causally related, given that the probability densities are correctly estimated. For this purpose, Tigramite offers three different methods: Gaussian Process and Distance Correlation (GPDC); a k nearest neighbor estimator (CMIknn); and an estimator based on kernel measures of CMI (RCOT). The author recommends using the most general conditional independence test, CMIknn, where multiplicative noise is expected, as in hydrology (e.g., Rodriguez-Iturbe et al., 1991), and where the sample size is smaller than 1000, which is the case in this study. CMIknn is, therefore, considered. The latter relies on a nearest-neighbor CMI estimator (Frenzel and Pompe, 2007; Vejmelka and Paluš, 2008) combined with a local permutation scheme as a nonparametric test for conditional independence (see Runge, 2018b).

Besides the maximum lag d_{max}, PCMCI requires other arguments. The PC stage retrieves parents according to a regularization parameter α_{PC} ranging between 0 and 1. The higher α_{PC}, the higher the number of parents, with $\alpha_{PC} = 1$ corresponding to the FullCI algorithm. If α_{PC} is too low, true parents might be missing. If α_{PC} is too high, the MCI step may retrieve spurious results due to the curse of dimensionality. When the ParCorr conditional independence test is selected, Tigramite allows optimizing α_{PC} while minimizing the Akaike Information Criterion (AIC, Eq.5.5). In this chapter, this feature is used to generate the ParCorr DAG. For CMI, the recommended values of $\alpha_{PC} = 0.2$ will be used (Runge et al., 2019a). The CMIknn further relies on a parameter defining the size of the neighborhood, k_{CMI}. The latter mostly act as a smoothing parameter regarding the CMI, and should not be too small (Runge, 2018b; Runge et al., 2019a). A final significance could be

adjusted to control the number of positive links. This chapter considers the standard significance threshold for the p-values: 0.05, 0.01, and 0.001.

In addition to the delay d_{max}, and the hypothesis of causal sufficiency (i.e., all common cause should be included), PCMCI relies on other hypotheses (discussed in Runge, 2018a): faithfulness, the Causal Markov condition; the absence of contemporaneous causal effects; stationarity; and the hypothesis and parameters of the underlying independent test. Faithfulness means that independence faithfully represents the absence of causality. Since PCMCI relies on conditional independence, some noise should remain while conditioning on the parents. Hence, as a consequence of faithfulness, time-series should not be strictly deterministically related to their parents. The causal Markov condition implies that conditioning on the parents of a variable makes it independent of all variables, including the parents of its parents (i.e., parents are sufficient causes).

The absence of contemporaneous effects is necessary to infer the direction of causality in virtue of the principle of priority. However, Tigramite allows reporting the contemporaneous dependencies. Similarly to CCF and CCM, this study reports them as bidirected straight arrows in the DAG. Regarding stationarity, it should be regarded not as the stationarity of statistical moment and spectrum of individual time-series but, more generally, as Eq. 6.3 and 6.4 being true for all time indices t. This hypothesis could be violated, for instance, if a new hydrologically connected path has developed being the experiment due to erosion, change in the land cover, or geomorphology. More obviously, the hypothesis is violated by the presence of no data in the dataset. However, Tigramite allows for handling no data automatically. The algorithm dismisses all time slices of samples where missing values occur in any variable for all lags up to $2d_{max}$. This feature will considerably reduce the size of the dataset for the real case study in this chapter but ensures that the causal test is always applied for the same time slices to avoid bias.

6.2.2 Study Cases

6.2.2.1 Virtual Experiment

This section describes a virtual case in order to become accustomed to the methods. A low dimensional, intelligible, and virtual hydrological experiment reproduces the problem of Figure 6-1 with two disconnected hydrological reservoirs (Figure 6-4). The confounding variable is the effective precipitation P_{eff}, expressed in mm, which represents the net flux of precipitation P minus evapotranspiration ET. In this case, P_{eff} is the only variable obtained from real daily monitored data: precipitation data at the RCL and the Penman-Monteith potential evapotranspiration data (Figure 3-6).

The conditionally independent variables (same as X_t and Y_t in Figure 6-1) are the discharge Q_A and Q_B of two disconnected reservoirs A and B. Both reservoirs take as input a net inflow term I_A and I_B resulting from a linear transfer functions H_A and H_B (or unit hydrograph) applied to a noisy effective precipitation input. Adding some noise in the model is mandatory to solve causality based on the independence of residuals with the PCMCI method. A multiplicative noise term is preferred as hydrological variables are most often characterized by multiplicative noise (e.g., Rodriguez-Iturbe et al., 1991). Then, the inflow terms are expressed as:

$$I_R = H_R * \left(P_{eff} + \varepsilon_R P_{eff}\right) \qquad Eq.6.6$$

where R is the reservoir name (A or B), the $*$ symbol denotes the convolution. Each reservoir will integrate a $P_{eff} + \varepsilon_R P_{eff}$ amount of effective precipitation. The term ε_R is a Gaussian noise term, different for each reservoir but following the same normal law $\varepsilon_R = \mathcal{N}(0, \sigma^2)$ of zero mean and varying standard deviations σ. In the experiment, the standard deviation σ varies based by a noise level factor linearly ranging by steps of 0.05 between 0.05 and 0.25 times the standard deviation of P_{eff}. Multiplying ε_R by P_{eff} ensures that the difference in the reservoirs inflows is potentially higher for extreme effective precipitations, mostly high rainfall events. For reservoir A and B, the unit transfer functions are respectively $H_A = [0.7, 0.2, 0.1]$ and $H_B = [0.1, 0.8, 0.1]$. H_A and H_B are the 3-day length window of convolution applied forwardly to the noisy effective precipitation attached to each reservoir. Hence, the reservoir A mainly responds to P_{eff} one time-lag before B, which

introduces a risk of a spurious causal association from A to B, as in Figure 6-1.

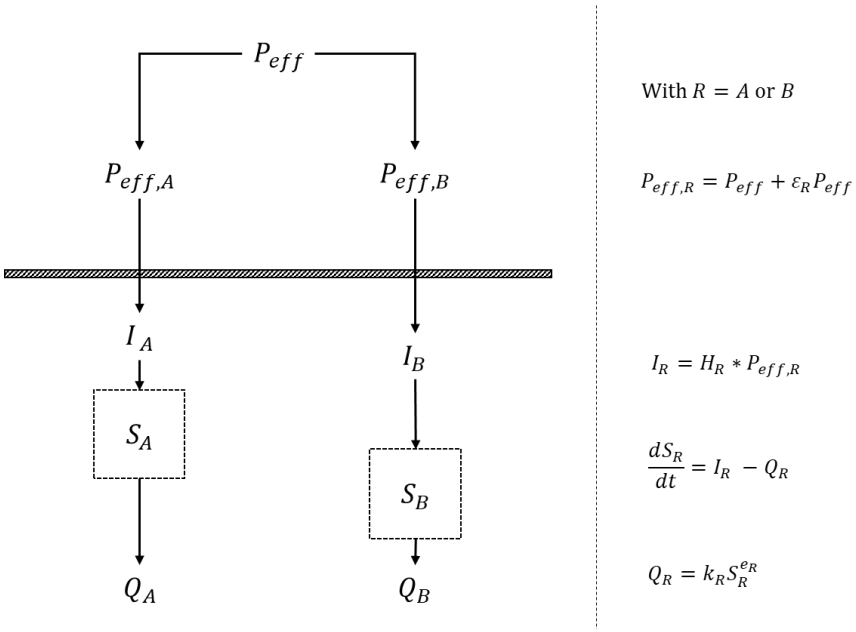

Figure 6-4: Conceptual and mathematical model for the virtual experiment. precipitation P_{eff} is the initial effective precipitation. With R the reservoir name, either A or B, $P_{eff,R}$ is the noisy effective precipitation for rerservoir R that is turned into an infiltration input I_R by applying a linear transfert function H_R. S_R is the time-variable reservoir storage, with an output discharge Q_R following a nonlinear storage-discharge relationship (Eq. 1.2).

Both reservoirs have a general nonlinear storage-discharge relationship $Q = kS^e$ (Eq. 1.2). Regarding the recession constant k and the nonlinear exponent e, three parametric scenarios are considered in Table 6-2. The first scenario considers the same parameters for the two reservoirs. The second one attributes a lower nonlinear exponent to reservoir B, which leads to a slower recession compared with the previous scenario. The last one set up a lower recession constant but exhibits faster recession due to a higher nonlinear exponent for reservoir B (see Figure 6-5).

Table 6-2: Parametric scenario for the virtual experiment.

Model: $Q = kS^e$	**Recession constant k**		**Nonlinear exponent e**	
Reservoir	A	B	A	B
Scenario 1	0.01	0.01	1.5	1.5
Scenario 2	0.01	0.01	1.5	1.2
Scenario 3	0.01	0.005	1.2	1.5

The synthetic dataset is obtained through numerical integration (using an Lsoda solver built in the *odint* function of the SciPy Python package) of the continuity equation $dS_R/dt = I_R - Q_R$, ensuring that $S_R \geq 0$. Within the model, the storage $S_R(t+1)$ is computed considering $I_R(t)$, which, therefore, introduces an additional lag between Q_R and P_{eff}. The model considers a warming-up stage of two years (2014, 2015) starting from initial conditions $S_A = S_B = 30$ mm. The final time-series have a length of 730 days corresponding to the years 2016 and 2017. Time-series for the three scenarios are shown in Figure 6-5 for the purely deterministic case of a noise level of 0. All the causal inference analysis considers the three variables P_{eff}, Q_A, and Q_B with a maximum causal delay d_{max} of 5 days, i.e., 1 step beyond the span of causal interactions. All time-series present intermittent periods where the discharge is null.

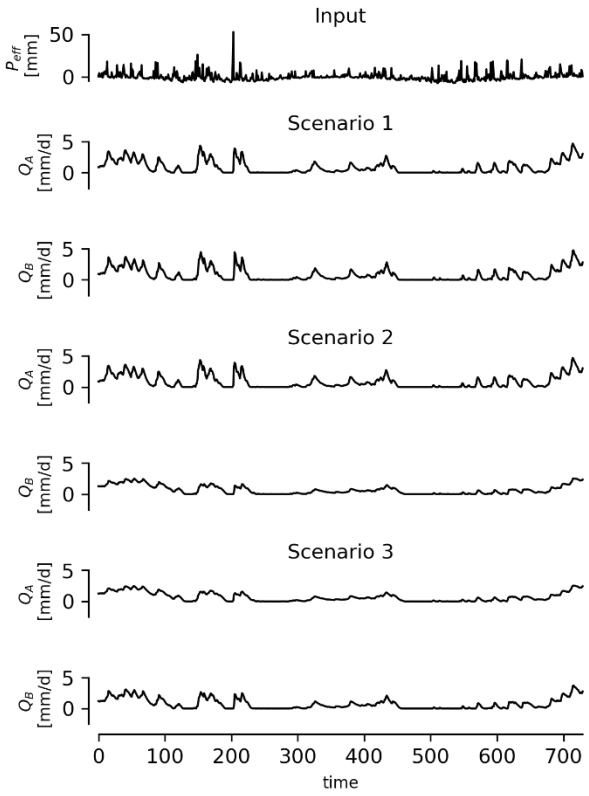

Figure 6-5: Synthetic flow series for the virtual experiment for each scenario in Table 6-2 without noise.

6.2.2.2 Hydrological Connectivity in the Vadose Zone

The first real application studies the time-dependencies between resistivity dynamical patterns and drip discharge data (P1, P2, P3, Figure 3-6). This is a case focused on hydrological connectivity, where we expect to reveal preferential flow paths with causal inference methods. Such a path or effective connection exists between P1 and the surface (Poulain et al., 2018; section 3.2).

Resistivity patterns are averaged time-series of the clusters shown in Figure 5-6.f (or Figure 5-3.h). The 6 clusters are the optimal clustering of raw resistivity series grouping linearly correlated series (considering *SI* in Figure 5-6.f). Their spatial distribution is similar to that of the groups obtained from the z-standardized log-resistivity data, the dynamics of which are shown in Figure 5-4 with a logarithmic scale. Since resistivity dynamics may not

include all common causes of percolation dynamics, precipitation and evapotranspiration data are included in the analysis. In doing so, the assumption of causal sufficiency is expected to be fulfilled for the multivariate causal analyses.

The maximum causal delay d_{max} is set to 5 days, covering the span of flood peaks in P1 (Poulain et al., 2018; section 3.2). The bivariate methods (CCF and CCM) are applied between all pairs of time-series. Many of these time-dependencies are expected to be causally spurious since the common causes are not taken into account. They will, however, indicate potential flow paths or functional connectivity.

Regarding the multivariate analysis, two scenarios are considered. The first one applies PCMCI without constraint on the parents, such that causality is solved entirely by the algorithm. The second one applies some constraint on the parent selection procedure: (1) ET and rainfall RF have no parents and are just considered as exogenous variables forcing the system; (2) resistivity series cannot have a drip discharge series as parents; (3) a drip discharge series cannot have another drip discharge series as parents (but its own past well). Hence such conditions impose a causal frame such that:

- internal variables of the system (resistivity) and their output (discharge) cannot influence the input driver (precipitation, evapotranspiration);
- internal variables (resistivity) can interact together;
- output (discharge) cannot interact with the internal variables. Otherwise, they would be input. Also, all outputs are causally independent as the product of its own system.

This scheme is consistent with conventional perceptual models of hydrological systems (e.g., Figure 1-1) and is considered in the eventuality that spurious DAG structure may arise from conditioning on an irrelevant variable (as suggested by Rinderer et al., 2018). Whether constrained or not, the time-slice considered in the analyses are the same. Hence, differences in the results cannot be imputed to the fact that the system is investigated at different periods.

6.2.2.3 Drivers of the Mass Balance

The second real application studies the system behavior holistically by including relative gravimetry measurements in the causal analysis (RG, Figure 3-6). P1 and P3 series are kept, but resistivity data and P2 are dropped due to their short time-domain and the high number of missing data. To consider all common causes of RG, the analysis includes potential evapotranspiration (EP), rainfall (RF), atmospheric pressure (RG). It is expected that P1 and P3 should not have much effect on gravity conditionally to the other variables except if their dynamics are representative of processes transferring water inside or outside the acquisition cone characterizing the scope of the gravimeter (section 3.3.3).

Similarly to the previous case, the delay will remain 5 days. The bivariate methods (CCF and CCM) are applied between all pairs of time-series, and the multivariate methods include a case where the parents are constrained. The conditions are: 1) ET, RF, and AP have no parents and are just considered as exogeneous meteorological variables forcing the system; 2) P1 and P3 do not have RG and GL as parents and cannot influence each other; 3) GL cannot be conditioned on RG. Hence, RG is the only variable that can have any parent.

6.3 Results

6.3.1 Virtual Experiment

Figure 6-6 reports the lagged dependencies regarding the spurious causal link $Q_A \rightarrow Q_B$ for all the causal inference methods applied to the synthetic dataset (Figure 6-5) with a multiplicative noise level factor of 0.15. A complete account of lagged dependencies is reported in Appendix VI (Figure VI.1 to VI.8).

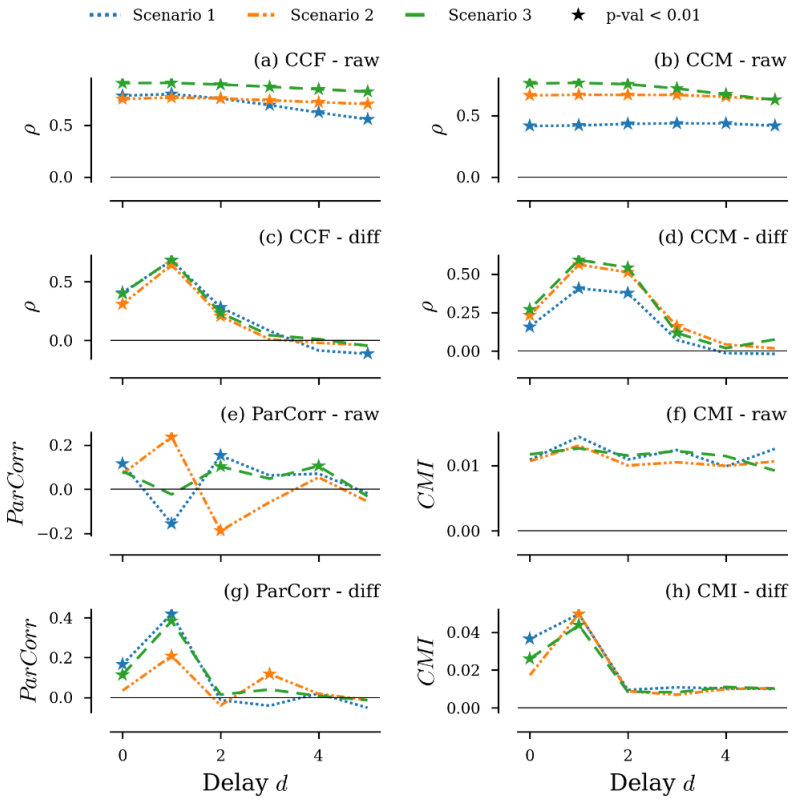

Figure 6-6: Time-dependencies measuring the delayed effect of Q_A on Q_B obtained with the four methods of Table 6-1 on the 3 scenarios of Table 6-2 considering a noise level factor of 0.15. The CMI method is applied with a k_{CMI} parameter of 20. Significant lagged dependencies are marked with a star symbol (p-val < 0.01). Both the raw data (raw) and their first order difference (diff) are considered to measure temporal dependencies.

Bivariate methods do not prevent the detection of significant, however spurious, causal associations between Q_A and Q_B. This is expected since bivariate approaches do not account for the principle of the common cause (d26 in Appendix III). On the raw data, CCF and CCM both present sustained time-dependencies either due to the shared seasonal patterns in the data, which may be seen as a common cause or to the long memory of reservoirs. These effects are removed while considering the differenced data (Figure 6-6, c and d), and both methods offer a better screening of the actual dependencies resulting from the model. CCM shows an additional significant dependence at a lag of 3 days that should not be considered (see Figure 6-3).

Regarding the multivariate approaches, only CMI applied to raw data does not detect any spurious relationship between Q_A and Q_B whatsoever the modeling scenario. Hence, considering the common cause P_{eff} allows retrieving a correct DAG structure. ParCorr does not. However, it could if the p-value threshold is adjusted to make absolute partial correlation non-significant below 0.2. Supplementary information (Appendix VI, Figure VI.9) shows that ParCorr is sensitive to the model's noise level when the noise factor is low. In such a case, the correct DAG structure cannot be retrieved with a higher significance threshold. The CMI method appears less sensitive to noise. However, a higher k_{CMI}, seems to identify the correct DAG slightly more often (see Appendix VI, Figure VI.11 and VI.12). Otherwise, increasing k_{CMI} has mostly the effect of smoothing and lowering the CMI value. These findings are in phase with what has been reported regarding the k_{CMI} parameter (Runge, 2018b; Runge et al., 2019a). As shown in Figure 6-6 (g and h), applying either the ParCorr or the CMI method on the differenced data never retrieves the correct DAG structure (Appendix VI, Figure VI.10, and VI.13).

As a result, bivariate methods (CCF, CCM) will be applied to the real dataset considering their first-order difference to better screen time-dependencies. Conversely, the multivariate methods (ParCorr, CMI) will be applied to the raw data. The k_{CMI} will be varied to verify the robustness of PCMCI-CMI outputs.

6.3.2 Hydrological Connectivity in the Vadose Zone

Figure 6-7 shows the bivariate linear dependencies obtained with the CCF methods. The CCF method is reporting many potential linear causal associations (see Appendix VI, Figure VI.14 for more details). If causality is hard to infer from such a diagram, the results make sense in general. A typical pattern is that the sign of time-dependencies tends to flip after a few delays. The reason is the common forcing of RF and the fact that dry periods come after the rain. Considering low delays, ET is positively related to the resistivity patterns, mostly at the surface (R1, R3). ET is negatively correlated with P1 only, which is known to drain fast flow from the surface through the epikarst. All variables are dependent on RF, the main confounding factor, but R0, associated with a dense limestone area, depends to a lesser extent. R4, the anomalous resistivity pattern, has a positive correlation. R0 and R4 put apart,

the quartet R1, R2, R3, and R5 exhibit strong positive and contemporaneous correlations together.

P1, P2, and P3 also are instantaneously related. P1 and P2 have strong dependencies with all resistivity patterns, but inconsistent and positive correlations are reported with the anomalous resistivity series R4. P3 seems rather dependents on R5 (slope), R2 (mostly limestone matrix), which makes sense since P3 is most likely draining the matrix's delayed flow.

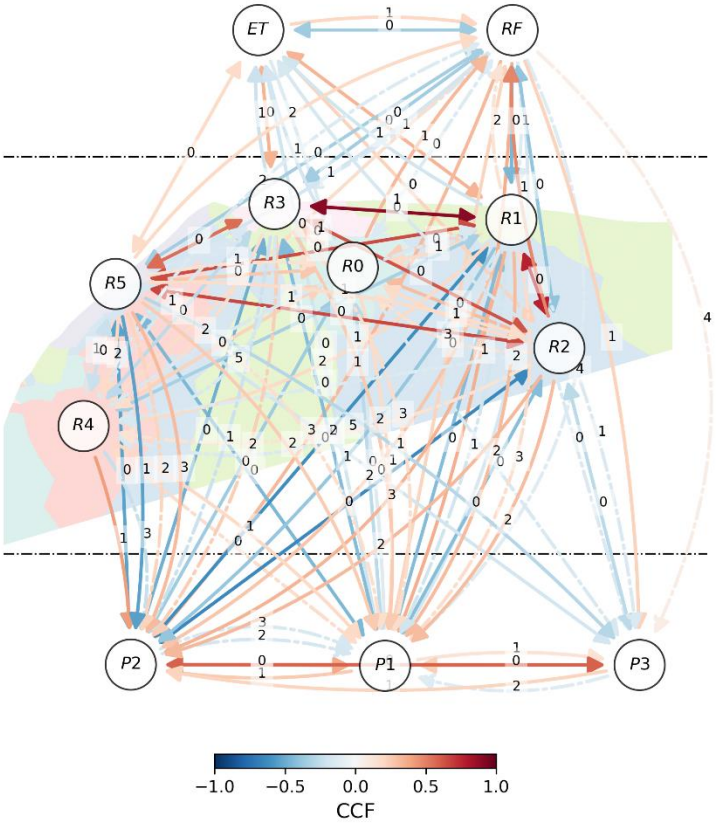

Figure 6-7: Graph of CCF cross-dependencies. Contemporaneous dependencies are represented by a bidirected straight arrow. Delayed dependencies are shown using directed curved arrows. All delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to CCF dependencies (Eq.6.2). Solid and dash-dotted arrows represent respectively significant dependencies with p-value < 0.001 and < 0.01. Variables are: evapotranspiration (ET), rainfall (RF), clustered resistivity time-series (R0: turquoise; R1: lime green; R2: blue; R3: pink; R4: red; R5: violet)), and drip discharge data in the Rochefort Cave (P1 to P3).

Figure 6-8 shows the bivariate nonlinear dependencies obtained with the CCM methods (see Appendix VI, Figure VI.15). Compared to CCF, the results are more intelligible since fewer links are reported. However, the nature of the dependencies is unknown and could be appreciated in view of CCF results (Figure 6-7).

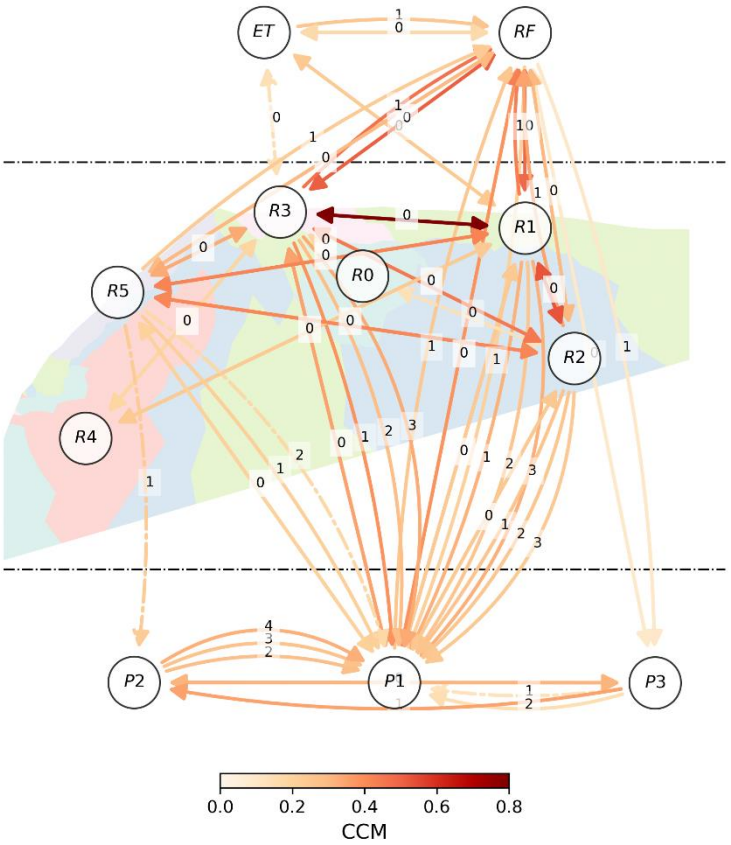

Figure 6-8: Graph of CCM cross-dependencies. Contemporaneous dependencies are represented by a bidirected straight arrow. Delayed dependencies are shown using directed curved arrows. All delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to CCM dependencies. Solid and dash-dotted arrows represent respectively significant dependencies with p-value < 0.001 and < 0.01. For a link to be reported in the DAG, the dependency on the next time delay must also be significant (see Figure 6-3). Variables are: evapotranspiration (ET), rainfall (RF), clustered resistivity time-series (R0: turquoise; R1: lime green; R2: blue; R3: pink; R4: red; R5: violet), and drip discharge data in the Rochefort Cave (P1 to P3).

With respect to drip discharge data, P2 is now exclusively related to R5, and P3 has no dependencies on resistivity patterns. P1 is CCM-related to R3 and the adjacent surface resistivity patterns R5, but also R2, which is somewhat representative of the limestone matrix resistivity. Compared to CCF, CCM supports the particular conclusion of preferential flows occurring between the surface and P1.

Similarly to CCM, ParCorr cross-dependencies (Figure 6-9.A) also reveal a significant link between R5 and P2, while P1 is associated with R1, R2, R3, and R5. The rainfall RF remains significantly related to P1, suggesting that resistivity patterns are not sufficient causes of P1. In Figure 6-9.A, no constraint is applied to the parents. Hence, for instance, ET or RF may have resistivity or drip discharge series as parents, which could be considered as irrelevant. Figure 6-9.B provides the results while a restriction is applied to some links. In doing so, P1 becomes only dependent on R1, and RF. Considering other clusters of resistivity, supplementary information (Figure VI.16) shows that P1 remains the main discharge series related to resistivity patterns when the DAG is unconstrained. With constraint, P1 remains related to one resistivity series; however, different in each case.

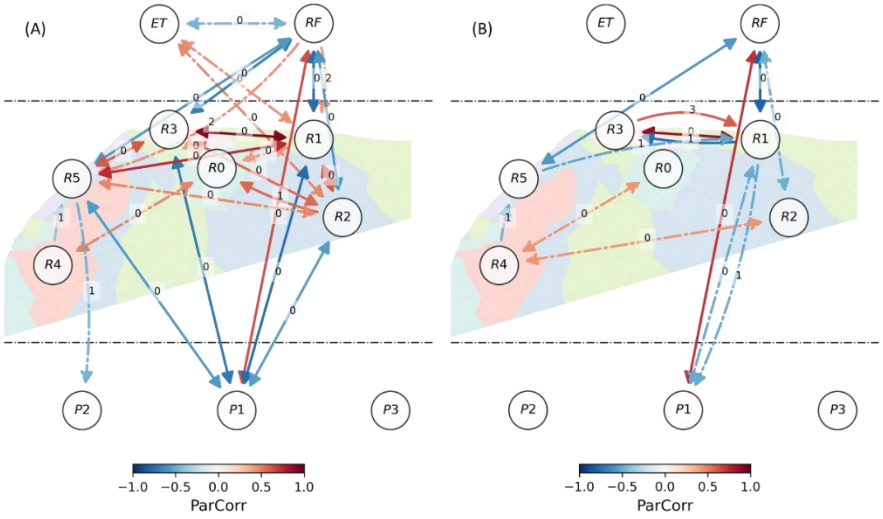

Figure 6-9: Graph of ParCorr cross-dependencies. (A) Unconstrained graph; and (B) constrained as follows: (1) ET and RF have no parents and are considered as exogenous variables; (2) resistivity series cannot have a drip discharge series as parents; (3) a drip discharge series cannot have drip discharge series as parents except itself. A bidirected straight arrow represents contemporaneous dependencies. Delayed dependencies are shown using directed curved arrows. All delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to ParCorr dependencies. Solid and dash-dotted arrows represent respectively significant dependencies with p-value < 0.001 and < 0.01. Variables are: evapotranspiration (ET), rainfall (RF), subsurface clustered resistivity time-series (R0: turquoise; R1: lime green; R2: blue; R3: pink; R4: red; R5: violet), and drip discharge data in the Rochefort Cave (P1 to P3).

Finally, Figure 6-10 shows the results for CMI independence tests considering a k_{CMI} of 20 (A.1 and B.1) and 25 (A.2 and B.2). A.1 and A.2 apply CMI without constraint on the parents, while B1 and B2 do. In contrast with the previous graphs, Figure 6-10 also includes dependencies with a significant level of 0.05 (dotted arrow). The unconstrained graphs (A.1 et A.2) present several arrows pointing upwards (e.g., $P3 \rightarrow R2$ or $P1 \rightarrow ET$ in A.1), which was not the case for the unconstrained ParCorr (Figure 6-9). Such upwards processes seem irrelevant. They do not appear in the constrained graphs (B.1 and B.2) since the constraint does not allow it. However, the irrelevancy of the results is further confirmed by the lack of robustness of the DAG while changing the k_{CMI} parameter from 20 to 25 (from A.1 to A.2 and B.1 to B.2). Further variation of the k_{CMI} confirmed the lack of robustness of CMI estimated with CMIknn on this dataset (Appendix VI, Figure VI.17).

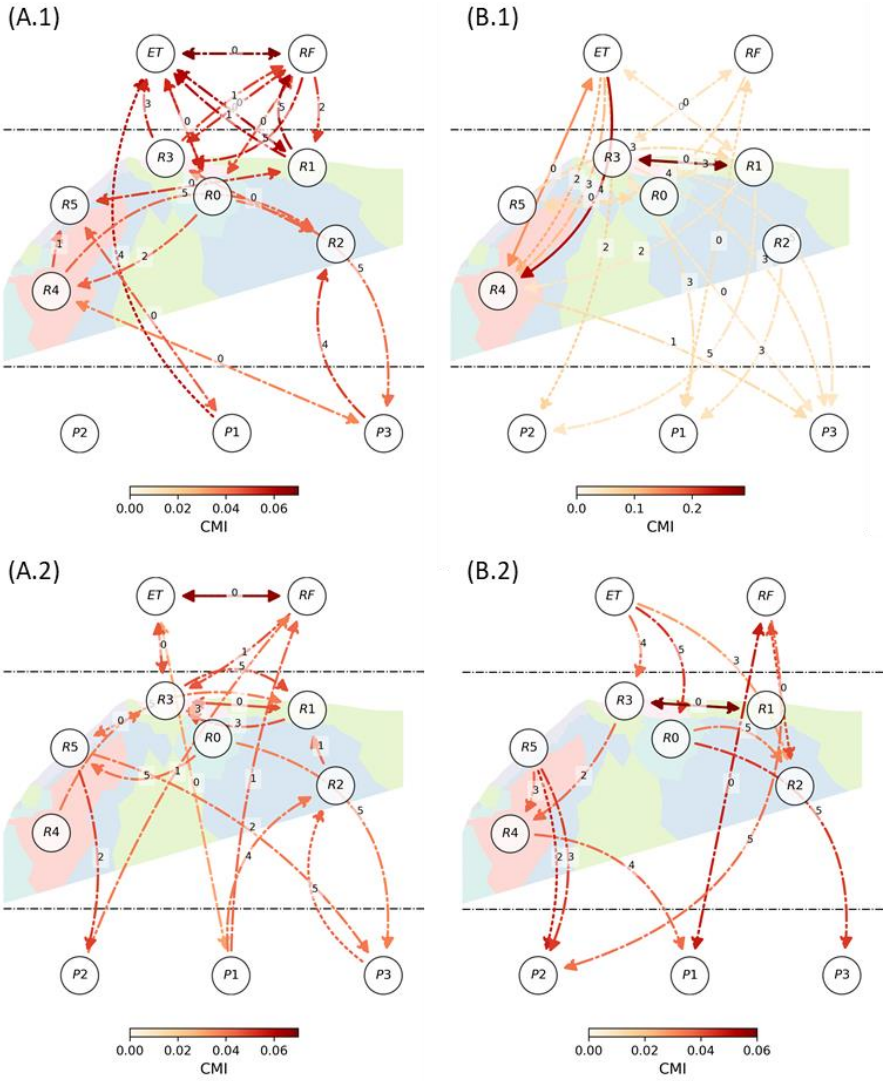

Figure 6-10: Graph of CMI cross-dependencies. (A) Unconstrained graph; and (B) constrained as follows: (1) ET and RF have no parents and are considered as exogenous variables; (2) resistivity series cannot have a drip discharge series as parents; (3) a drip discharge series cannot have drip discharge series as parents except itself. A.1 and B.1 uses a k_{CMI} of 20. A.2 and B.2 uses a k_{CMI} of 25. A bidirected straight arrow represents contemporaneous dependencies. Delayed dependencies are shown using directed curved arrows. All delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to CMI dependencies. Solid, dash-dotted, and dotted arrows represent respectively significant dependencies with p-value < 0.001, < 0.01, and < 0.05. Variables are: evapotranspiration (ET), rainfall (RF), subsurface clustered resistivity time-series (R0: turquoise; R1: lime green; R2: blue; R3: pink; R4: red; R5: violet), and drip discharge data in the Rochefort Cave (P1 to P3).

6.3.3 Drivers of the Mass Balance

Figure 6-11 shows the DAG obtained with the different methods and scenarios. With respect to CCF and CCM, the detailed time-dependencies are reported in Appendix VI (Figure VI.18, VI.19). Complementarily, Figure VI.20 shows the unconstrained and constrained CMI scenarios (Figure 6-11, E and F) for various k_{CMI}.

Linearly, RG is mostly affected positively by GL, instantaneously but also at lag 1. AP presents correlations with RG at lags 0 and 1, suggesting that the pressure effect is imperfectly removed. However, AP is negatively correlated with RF. Thus entirely removing the atmospheric effect by a linear regression will probably remove some information about rainfall. Rainfall presents a negative correlation at lag 0 and a positive one at lag 1. Such delay could be explained by the fact that RF influences GL mostly with a lag of 1 day, or potentially considering slow flows and the necessary time for the water to percolate down to the gravimeter's scope in the first few meters. RG has positive linear dependencies with P1 (lag 0 and 1) and P3 (lag 0). As it is observed in Figure 6-7, the sign of significant time-dependencies frequently switches after some duration.

CCM nonlinear dependencies usually do not consider these further lagged dependencies. Regarding RG, it is no longer dependent on P1. The lag-1 RG-RF relationship is no longer significant, and a closer look at Figure VI.19 suggests that it is RG causing RF. RG remains highly related to GL, and, to a lesser extent, to AP.

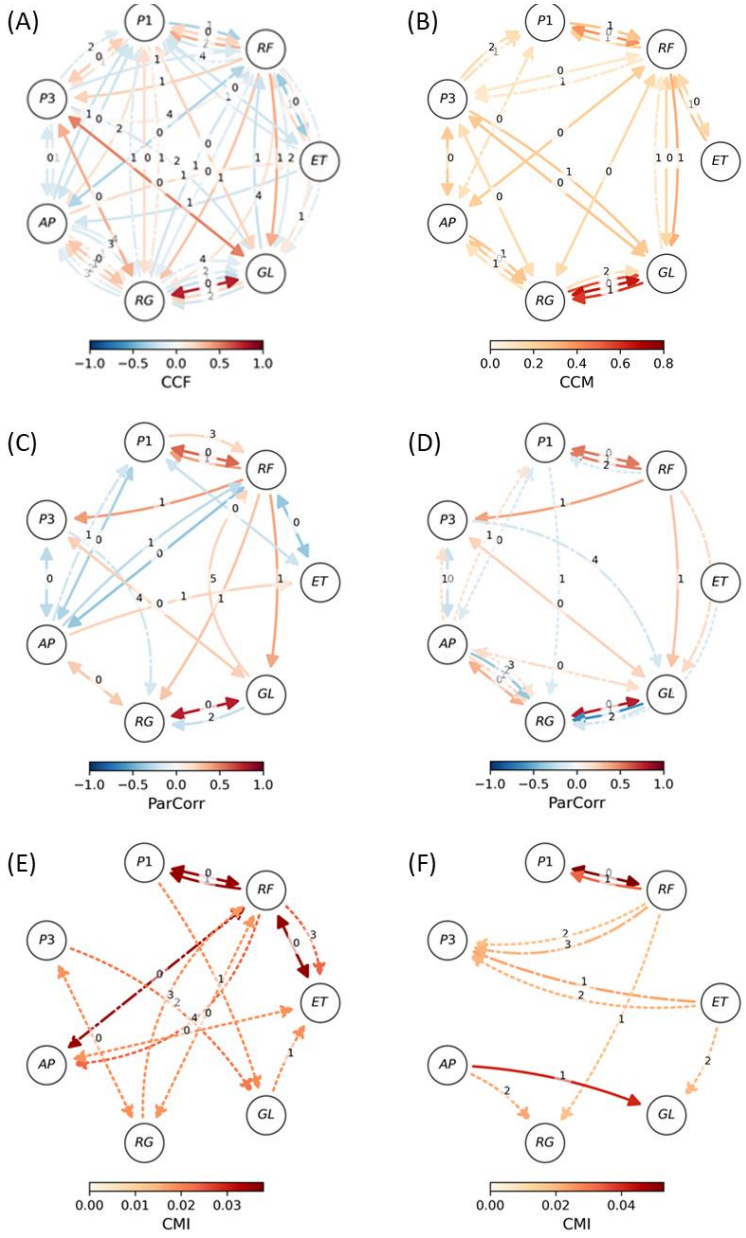

Figure 6-11: Graph of cross-dependencies. (A) CCF; (B) CCM; (C) Unconstrained PCMI-ParCorr; (D) Constrained PCMI-ParCorr; (E) Unconstrained CMI with $k_{CMI} = 30$; (F) Constrained CMI with $k_{CMI} = 30$. Variables are: evapotranspiration (ET), rainfall (RF), groundwater level (GL), relative gravimetry (RG), atmospheric pressure (AP), and drip discharge data (P1, P3). A bidirected straight arrow represents contemporaneous dependencies, curved arrows for delayed. All delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to the color bar scaling the time-dependencies. Solid, dash-dotted, and dotted arrows represent respectively significant dependencies with p-value < 0.001, < 0.01 (and < 0.05 for D, E, and F only).

Regarding ParCorr dependencies, most of the links are consistent with those revealed by CCM. They are also consistent between the constrained and unconstrained analysis (C and D), considering that some links cannot occur in the constrained graph (e.g., between meteorological variables, ET, RF, and AP). The unconstrained graph presents a few strange links, such as P1 causing RF at lag 3 or GL at lag 5. Regarding RG, it is still related to GL and AP. The direct influence of RF on RG disappears in the constrained graph (D). The drip discharge data negatively impact gravity (P3 at lag 4 in C, P1 at lag 4 in D) and potentially represent some flows leaving the system.

Regarding CMI, the graphs on this longer dataset seem less erratic than those obtained in the previous study case. Nevertheless, they remain sensitive to k_{CMI}, especially regarding the link of low significance, that sometimes are found strange and spurious (e.g., RG→RF, GL→ET in Figure 6-11 E). On the unconstrained graphs, some links are systematically obtained. CMI detects the link between P1 and RF, which was not the case in the previous section, and further encourage the idea of P1 draining fast preferential flows. Links between RF and ET, as well as RF and AP, are stable. However, no link seems stable regarding RG. This is not the case on the constrained graph, and AP is the only variable that causes RG consistently. CMI does not detect the strong relationship that is supported by the other methods between RG and GL.

In general, RG responds to GL, AP, and RF. P1 and P3 have a positive influence that disappears while considering ParCorr. This is consistent since they are internal variables of the system, and water transfer within the gravimeter's scope should have a small impact on gravity variations.

6.4 Discussions

6.4.1 On the Robustness of CMI

An important question is: why was the most promising PCMCI-CMI method the less robust method on the real study cases? A possible answer is the lack of data, since excluding the short time-series of P2 and resistivities provided robust significant links. Another possibility is the nature of data, resistivity series comes from a smooth inverted model, and the method would be sensitive due to the difficulties of extracting resistivity patterns (Chapter 5).

Despite the sensitivity of the method to the k_{CMI} parameter, it is also possible to consider that some values are preferable to others from a hydrological point of view. The aim of k_{CMI}, which define the size of the neighborhood, is to capture the state of the systems in a meaningful way into probability densities (Eq. 6.5). The densities are most likely affected, for a hydrological system, by the rainfall patterns in terms of frequency, intensity, duration of rainfall, and conversely, the length of dry periods. Therefore, a meaningful k_{CMI} would be dependent to the time-resolution and would be case-specific. More insight may come from further virtual experiments conducted, for instance, with a stochastic rainfall generator allowing to generate many rainfall patterns and relate them to the selection of k_{CMI} and the necessary length of the dataset to ensure meaningful outcomes. However, as CMI has high computational requirements using a nearest-neighbor estimator (CMIknn), this should be conducted on a high-performance computer. With the same computing facilities, it would also be possible to select a DAG based on cross-validation or model evaluation, as it is done for ParCorr. Indeed, such DAGs are models, not just graphical models.

6.4.2 Causality, Physics and Effective Connectivity

Empirical causality is inferred from dynamics, and it is essential to remind that such an approach gives no physical meaning to the variable quantities. For instance, some links between P1, P3, and RG or GL have been reported. Such links cannot be interpreted as physical causation as a few litters of water may not influence RG or GL. Empirically, they may, because their dynamics are potentially representative of processes occurring at a larger scale. Consequently, it is delicate to claim effective hydrological connectivity between the two monitored dynamics. In other terms, connectivity patterns by such methods may remain attached to functional, while being more likely related to effective connectivity, following Rinderer et al. (2018) terminology.

That being said, it is the practitioner's responsibility to advocate for a functional and representative causal relationship representing the system's general behavior or for an effective connection. In our case, empirical causality was consistent with the dye tracing test done in situ and the geological features of the site (Poulain et al., 2018; Watlet et al., 2018b). Hence, effective connectivity between the surface and P1 is the best hypothesis, but not obtained from the single scope of time-series analysis. The advantage of time-series analysis over dye tracing is its ability to study the

dynamics over a longer time-span and the time-variant and state-dependent connectivity patterns. Hence, a perspective is to apply causal inference between different seasons or antecedent hydrological conditions and compare the retrieved patterns to test if effective connectivity detected with dye tracing remains while the system is at different stages. This perspective is related to the question: should DAG be static or dynamic? In general, the effect of intermittent dynamics, which are intermittent causal relationships, was not covered in this chapter. The virtual experiment suggests that it is possible to apply these methods when processes are intermittent, and real cases have included daily precipitation data.

6.4.3 On Potentially Spurious Causal Associations

The previous section reports a first cause of spurious causal associations, which would be semantic, i.e., mistaking a representative and functional causal association with an effective and physical causal association, such as the one revealed by a tracer. Regarding effective connectivity, retrieving causality empirically from a hydrological system is problematic due to the general forcing of meteorological variables. Most variables respond synchronously to these inputs such that the direction of causality cannot be inferred from the principle of priority of cause (see zero lag bidirected straight arrows in all graphs). Regarding asynchronous time-dependencies, they tend to be ubiquitous and thus motivates the reliance on state dependencies methods such as CCM or the use of a multivariate framework for causal discovery for the focus they bring on a reduced set of dependencies.

Otherwise, in a multivariate framework, spurious causal associations may arise from violating the hypotheses underlying the causal inference method. In this case (Runge, 2018a), they are causal sufficiency, faithfulness, the causal Markov condition, the absence of contemporaneous dependencies, stationarity, and hypothesis and user-defined parameters behinds the independence test (section 6.2.1.3). Bivariate methods do not consider the principle of the common cause, but multivariate methods may not fulfill causal sufficiency. Fulfilling causal sufficiency is not trivial. Usually, hydrological data are either missing or high dimensional and remotely sensed spatiotemporal datasets. As the ERT dataset, such datasets should priorly be dimensionally reduced into a few representative variables, which open the perspective to study causal inference method in parallel of dimension reduction methods (e.g., Chapter 5; Appendix VI, Figure VI.16). More

generally, modeling artifacts, when data are inferred or corrected with a model, could result in spurious links, as it may be the case with the ERT dataset (see, for instance, causal dipoles forming in Appendix VI, Figure VI.17, A.2, A.3). Another possibility is that such time-series are too deterministically related and violate the faithfulness hypothesis when a nonlinear test is considered. When different sensors directly measure data, they may have different precisions and error models, either driven by dynamic or instrumental noise. Noise structure is expected to interfere with the causal inference problem since conditioning rapidly removes shared trends and harmonics in the signal. As a result, one could work on noise reduction or noise harmonization between the time-series to make the analysis more robust, keeping in mind that noise is mandatory in a conditional independence framework.

Nonstationarity is also a potential cause of spurious DAG. Nonstationarity (i.e., Eq. 6.3 and 6.4 being true for all time indices t) could be the result of failing to causal sufficiency (Runge, 2018a). Problematically, nonstationarity could be the result of land cover or geomorphological change. Such structural changes are not easily tractable and turned into a time-series to be included in the analysis. In that regard, testing nonstationarity on individual time-series with nonlinear state space methods for cross-prediction allows us to test if the state dependencies remain consistent through time (Schreiber, 1997; see also 4.4.1).

6.4.4 On Constraining Causality

Spurious associations may negatively affect the results by conditioning on irrelevant parents. For this reason, it was suggested to constrain the causal parents in order to increase the reliability of causal graphs. However, this task is delicate. First, the method loses a part of its empirical status by forcing some hypotheses. Secondly, while section 6.4.3 accounts for the reasons for true spuriousness, it is above all the experimenter's judgment that qualifies a relationship as spurious. If this judgment is erroneous, constraining would be counter-productive. There is also a risk of falling into a similar version of the paradox of inquiry: if causality is known, there is no need to discover it; if causality is not known, there are no clues as to where and how to search.

If one considers constraining, the task is to advocate for conceptual models of the system that are justified. In this case, the meteorological forcing variables

(rainfall, potential evapotranspiration, atmospheric pressure) were considered as causally independent variables to better depict their direct effects on each internal variable and output. They could not be conditioned on the internal variable of the system as well. Nonetheless, some studies consider a causal effect of soil moisture on precipitation (Salvucci et al., 2002; Tuttle and Salvucci, 2017; Wang et al., 2018), which is legitimate considering the hydrological cycle. Again, this is a question of representativity versus what is physically possible. In our case, resistivity patterns at the surface were causing precipitation with the CCM method. They may indeed be functionally representative of soil moisture patterns at a larger scale, possibly influencing rainfall patterns when precipitation is convective, but certainly not as a physical cause. Arguably, it is neither a direct cause of rainfall but a transitive one through its influence on real evapotranspiration.

Outputs as drip discharges in the first real experiment (section 6.2.2.2) were causally independent but remained conditioned on their past values. The past self-conditioning of drip discharge data makes sense since they are the physical outputs of reservoirs exhibiting autoregressive patterns. However, the past self-conditioning of gravity raises more questions, depending on how the system is framed. If one thinks of relative gravimetry (RG) as an eye on the hydrological mass balance, it is not supposed to cause itself physically. The autoregressive pattern of gravity is instead imputed to the autoregressive behavior of the monitored water reservoirs.

Besides, the theory of nonlinear dynamical system tells us that all the causes are embedded in the trajectory of one single variable (section 4.2.2.1). Hence, conditioning gravity on its past potentially hides the effect of other causes embedded in its own trajectory. However, not to do so would have been to neglect Wiener-Granger's definition of causality, i.e., causes are supposed to bring more information on the trajectory of a variable than its past. Granger was an econometrician, a field where it makes sense. Financial time-series have an autoregressive and self-driven behavior, with growth favoring growth and recession favoring recession, however responding to exogenous shocks. Wiener was the father of cybernetics, which relates to the study of navigation systems. Controlled navigation is following a self-driven trajectory that also responds to new information as exogenous shocks. The two researchers certainly had little concern about connectivity and the possibility of physical interactions since stock markets guarantee the interactions between the variables, as sensors and electronic circuits regarding cybernetics. Knowing

this, hydrologists have the opportunity to develop and explore alternative definitions of causality that are more appropriate to their purpose and discipline (section 2.5).

In practice, the constraint did not make much difference when using ParCorr in the first experiment (section 6.2.2.2) and did not affect the instability of CMIknn. In the second real experiment (6.2.2.3), the constraint has the effect of revealing time-dependencies between RG and the atmospheric pressure AP both using ParCorr and CMIknn with various k_{CMI}. Such delayed causal effect can be hypothetically explained as a delayed effect of atmospheric pressure on Earth crust or as the result of spatiotemporal patterns of atmospheric pressure at a larger scale since the AP signal provide a local correction while the gravimeter capture atmospheric pressure on a broader area (see Watlet et al., 2020).

6.5 Conclusion

Four causal inference methods were tested on hydrological datasets. The first two are bivariate methods: the Cross-Correlation Function (CCF) and Convergent Cross Mapping (CCM), analyzing respectively linear and nonlinear time-dependencies between two variables. The last two are multivariate, and both involve the PCMCI algorithm with different conditional independence tests: the linear partial correlation test (ParCorr) based on a multivariate linear model and the nonlinear Conditional Mutual Information (CMI) test using a nearest-neighbor estimator of CMI. Three study cases were proposed: (1) a virtual experiment with two parallel nonlinear hydrological reservoirs fed by a linear transfer function applied to effective precipitation; (2) a case study involving potential evapotranspiration, rainfall, clustered resistivity patterns in the subsurface above the Rochefort Cave, and drip discharge in the Rochefort Cave; and (3) a similar dataset, not considering resistivity patterns, but atmospheric pressure data, groundwater level, and relative gravimetry data.

The virtual experiment showed that bivariate methods are best applied on differenced data to identify causal time-dependencies. Bivariate methods, however, cannot make the difference between functional connectivity due to the common forcing of reservoirs and effective connectivity. On the other hand, multivariate methods could make this difference by considering the raw (i.e., not differenced) data. Compared to ParCorr, PCMCI-CMI is

theoretically the most reliable test to retrieve effective causal interactions empirically.

On the two real study cases, however, CMI provided less robust and interpretable results. CCF renders linear associations among variables and is therefore interpretable, but these associations are ubiquitous, and therefore, no preferential flow or general functioning of the systems arises. CCM and PCMCI-ParCorr matched well with each other's and allowed identifying P1 has a preferential flow path and explaining the relative gravimetry measurements. ParCorr did not output many irrelevant interactions while being unconstrained. Showing much lesser interactions allows focusing on important relationships in the system and is, therefore, more informative than the usual CCF method. In particular, the discussions cover four aspects: (1) the problem of the robustness of the CMI method applied to real data; (2) the need to discuss and confront empirical causality with physical causation and tangible proofs of effective connectivity; (3) the potential reasons for the occurrence of spurious causal links; (4) the conceptual problems of constraining and forcing variables that are causal or not in the analysis.

In real cases, there is always a risk of spurious causal associations, i.e., a functional link but not an effective one, because the hypotheses behind the methods cannot be adequately controlled. Since results vary substantially across the datasets, methods, and hypotheses, single p-values do not relate to the probability of causal associations. Concerning causal inference methods, their number is progressively increasing, input data can be selected and arranged in many ways, the causal graph possibly constrained following different logics, and some methods may be sensitive depending on the study case. Accordingly, the use of one single method reporting significant associations is discouraged, especially without sensitivity analysis, the testing of various scenarios, the introspection of the underlying hypothesis of the causal inference method, and interpretations on the conceptual framing of the system and its potential causal mechanisms.

Chapter 7 General conclusion

"Reading furnishes the mind only with materials of knowledge; it is thinking that makes what we read ours."

John Locke

The initial intention behind the Ph.D. project was to capitalize on the recent advances in time-series analysis to draw a better understanding of karst systems, in particular, the Lhomme Karst System (LKS), in Belgium (section 1.2.1, Chapter 3). In practice, the thesis capitalized on the recent advances in the understanding of the LKS (Poulain, 2017; Watlet, 2017) to answer the question: what does it take to understand complex hydrological systems from empirical time-series analysis?

To that regard, the introduction postulates that complexity first depends on the observer, his or her ability to observe, and how he or she seeks to interpret a hydrological system (section 1.1 and 1.1.2.1). Complexity can also depend on the system's inherent characteristics, but that these characteristics are difficult to separate from the way the observer sees things. These are dimensionality, i.e., the number of variables driving the system (section 1.1.2.2), the nature of the mechanisms between the variables (section 1.1.2.3), and organization, which gives an idea of how the variables and mechanisms are ordered as a whole (section 1.1.2.4). From these points of view, a system could be considered complex in a non-mutually exclusive way if it has high dimensionality, presents nonlinear mechanics (sensitive dependence to initial conditions), and is disorganized. The thesis explores these four topics in its chapters.

7.1 What Does it Take to Understand?

Chapter 2 explores the general question of what does it take to understand? The question is addressed by portraying the historical evolution of the concept of causality and mirroring it to epistemological concerns or philosophies in the science of hydrology. Undoubtedly, history has given us many perspectives on how to estimate what is true or not, causal or not, since the chosen starting point in Ancient Greece (section 2.3). Causality is subject to

many concerns: is it a perception, a true belief, and, if true, to which account (section 2.3.2)?; how to deal with causality with the problem of nonstationarity or change (section 2.3.3.3), which relates to the problem of induction (section 2.4.3.1)?; the concerns related to the problem of universals and the realism of models (section 2.4.1). Different frameworks have been proposed to deal with causality: the initial one of Aristotle (section 2.3.3), the rationalist framework focused on deduction and mechanism (section 2.4.2), the empirical one inferring causality from data (section 2.4.3), the perceptual one inferring causality from people (section 2.4.4), and the systemic one that makes a synthesis of the former causal frameworks that goes back to Aristotle and emphasizes the importance of the whole (section 2.4.5).

Chapter 2 concludes with a unifying philosophical framework that suggests that what is causal arises from the convergence of these perspectives into a stable community consensus constrained by logic (section 2.5). Causality is not an abstract product of our imagination nor reality. Causality is resulting from the interactions of our concepts and logic with the real world, which allows it to be realistic without being real or purely abstract. This philosophy is a vital and positive state of mind allowing proactive research on causality since it suggests that not all models are wrong (in contrast with Box, 1976) if they aim at causality instead of reality. Causality aims at understanding, so the usefulness of a model is not sufficient to call it right, in the sense of causal, but necessary as it has to prove its connection to the real world. To be completely causal, the nature of a causal explanation, on which there is consensus, must also take into account how the system is conceptualized. Conceptualization involves making this judgment about the elements to be taken into account (solving dimensionality), the processes that relate them together (solving mechanisms), to form a whole system (solving organization), for a given context and purpose that provides a frame, inherently limited, but necessary to justify and evaluate the proposed causality. This view is close to the systemic and Aristotelean philosophy of causality.

The daily objective of a scientist is not always to provide causal explanations, for example, with causal models, but more often to provide elements of explanations in a narrower framework that can contribute to the construction of causal explanations at the community level. It is in this spirit that the following chapters contribute to causality.

7.2 Causality from Nonlinear Patterns

Chapter 4 develops a method that relies on empirical (section 2.4.3) and univariate time-series analysis to assess river recession dynamics' complexity, with complexity attached to nonlinearity (sections 1.1.2.3 and 4.1). Since the river hydrograph is the output of a catchment system, the approach is systemic (section 2.4.5), as it is assumed that complexity is a trait of the whole system. The proposed method, EDM-Simplex (section 4.2.1), is a nonlinear empirical model that forecasts recession segments' future from the future of similar recession segments. The decay of the forecasting performance with the prediction horizon (1, 2, and 3 days ahead), while remaining very good (Figure 4-6), was found to be a robust indicator of an empirical definition of nonlinearity, that is, sensitive dependence to initial conditions. The method was tested on three hydrological series from gauging stations located before, inside, and after the karst system (S1, S2, S3 in Figure 3-1 and Figure 3-3). EDM-Simplex allows showing that the nonlinearity of recession tends to increase from upstream to downstream due to the karst system (Figure 4-5). This increasing trend is not obtained with the conventional method based on the parametric model of Brutsaert and Niebert (B&N, Eq. 1.5), where nonlinearities were ranked as S2 < S1 < S3 (Figure 4-2).

As a result, the study shows the importance of framing the analysis with methods and definitions, and most valuably, in considering different ways to do it. The B&N is not wrong nor less suited; what it captures is a trend. Hence, it looks at the dynamic at a different scale, which inevitably changes causality. The percolation through the karstic riverbed is the dominant process at S2. It can explain the relative linearity of S2 retrieved with the B&N method. However, trends do not capture the dynamic in detail. EDM-Simplex does and reveals that complexity is increasing, which is valuable information for modelers looking for a fit representation of the system, further than dominant processes. In that sense, EDM-Simplex is superior to B&N but does not allow identifying trends.

Also, depending on the definition of recession points or segments, the EDM-Simplex results may change, which is not a sign of lack of robustness, but the simple evidence that a system does not have the same sensitivity according to its states. For instance, nonlinearity was lower for the downstream station of Eprave (S3) compared to the nonlinearity within the karst (S2) when EDM-Simplex forecast recession points representative of low flow (Figure 4-7).

This result highlights the importance of the Vauclusian spring during low flow, and that S3 is therefore not following the patterns of upstream recession. All these examples support the view that the full conceptualization of a system arises from combining methods, definitions, and purposes. A global understanding and consensus arise when particular behaviors have found an explanation in their context, including our prior conceptual understanding of the system (section 2.5). If empirical and systemic approaches do not highlight the exact processes involved in causality, such experiments suggest their implicit existence, a necessary step before their explicit representation.

7.3 Causality from Dynamic Similarity and Dimension Reduction

Chapter 5 investigates causality or "complexity reduction" through the dimension reduction and the extraction of spatiotemporal patterns (hydrofacies) from an Electrical Resistivity Tomography (ERT) model of the subsurface (section 3.3.2) at the Rochefort Cave Laboratory (RCL, section 3.2). Methods of time-series clustering (TSC) are increasingly popular for this task, which relates to causality from three points of view. First, the high dimensionality of ERT model, a spatial mesh of 1558 cells times 465 days, in this case, makes it difficult to understand (sections 1.1.2.2 and 2.4.2). Secondly, the exercise of extracting emergent patterns relates to the systemic perspective of investigating the forms of organization (sections 1.1.2.4 and 2.4.5). Thirdly, TSC is mostly empirical (section 2.4.3), given that the clustering involves averaging time-series together based on their dynamic similarity (i.e., correlation). The dynamic similarity is what possibly and empirically defines hydrofacies, that are, from the perceptual point of view, lithological features showing subject to similar hydrological forcing and processes. Confronting both definitions allows judging if the clustering makes sense from the perceptual understanding and physical knowledge of the system (section 2.4.4, Figure 3-5). Chapter 5 examines various grouping algorithms from the literature, time-series representations, and another subtle nuance in the definition of hydrofacies: whether or not they should be spatially contiguous since contiguity is a feature of causality (d23 in Appendix III).

Inevitably, the results show differences between the combinations (e.g., Figure 5-6) since each approach maps to another definition of hydrofacies. However, some patterns are redundant and in phase with the perceptual model, which leads us to consider them consensually as hydrofacies. However, the

patterns obtained from the full dataset are not retrievable when TSC is applied to shorter time samples (Figure 5-7). This result shows that hydrofacies are particularly hard to generalize for small ERT time-spans, i.e., the problem of induction (section 2.4.3.1). Technically, it is possible to define a clustering method based on consensus among various approaches, which could fasten the convergence to a stable model. Another option is to consider other geophysical datasets in the clustering.

In the end, Chapter 5 is a study that promotes understanding of TSC methods and awareness of their limitations, starting from a very generic definition of hydrofacies, which provides broad guidance to practitioners. As is the case with causality, clearer guidance can be given to goal-oriented studies on a particular hydrological or geophysical purpose, not for the only sake of clustering. In such cases, the definition of hydrofacies takes on a more precise meaning and offers a better roadmap. For example, in Chapter 6, which relies on a clustered ERT dataset, the spatial constraint of contiguity was not preferred over the idea of having groups that are as uncorrelated as possible. If hydrofacies should exist more generically as a universal concept (section 2.4.1), it is up to the community to refine its definition by asking, for example, some relevant and interrelated questions: why hydrofacies are useful as such?; what patterns characterize them? in which domain of the ERT signal spectrum?; how they emerge from the mechanism of hydrological processes?; should they be contiguous, and are they allowed to change over time? Answering these questions paves the way to the method, as a way to frame causality.

7.4 Causality from Time-Dependencies

Chapter 6 assesses the ability of causal inference methods to remove complexity by inferring from times-series the organization of a system in the form of a causal graph, or Directed Acyclic Graph (DAG). The approaches are empirical (section 2.4.3). Significant time-dependencies could be seen as Hume's principle of constant conjunction (d24 in Appendix III). Besides, the methods rely either or both on Hume's principle of priority of the cause over its effect (d22) and the principle of the common cause (d26) to dismiss dependencies resulting from a shared driver. The approach is also systemic (section 2.4.5) since causal inference methods should be applied on a set of hydrological series representative of all the dominant spatiotemporal patterns

found in the system. The DAG reports the temporal links between variables, a way to depicts the organization of the system.

If causality remains widely assessed with the cross-correlation function (CCF), Chapter 6 shows that the CCF method produces ubiquitous causal links in a hydrological system due to the general forcing of meteorological variables. CCF thus depicts a complex system, high dimensional (section 1.1.2.2), and disorganized (section 1.1.2.4), that is, therefore, difficult to understand (section 1.1.2.1), but still interpretable because the method is linear (section 1.1.2.3). The Convergent Cross-Mapping bivariate method (CCM, section 6.2.1.2) that looks for nonlinear state-dependencies between two variables pictures more parsimonious DAG with fewer links. They are more understandable in that sense while being less interpretable on the nature of the dependence as a nonlinear method. Encouragingly, the CCM supports hydrological connections and preferential flow between the surface and percolation in Rochefort's cave, where such a link has been revealed by a dye tracing test. However, spurious links could arise since CCM does not consider the principle of the common cause.

The multivariate framework tests for conditional independence (PCMCI, section 6.2.1.3) and takes into account the principle of the common cause, using either the partial correlation (ParCorr) or the conditional mutual information (CMI), respectively, as a linear and nonlinear test. ParCorr also supports the occurrence of preferential flow in the cave while showing some differences with CCM. In contrast, the CMI test does not provide robust results and interpretable links in the real case studies that were considered. The structure of the DAGs varies erratically, depending on the parameters chosen to estimate the CMI, and produces too many alternatives to judge them understandable and trustable (section 1.1.2.1, e.g., Figure 1-2). Possibly, this unsatisfying outcome may be due to a lack of data or modeling artifact. However, CMI is theoretically the most robust method as it takes into account the common cause and does not make a hypothesis on the mechanism, whether linear or not, and its robustness was verified in a virtual hydrological experiment (section 6.2.2.1).

In general, nonstationarity, missing important variables, the lack of data in quantity or quality, noise structure, modeling artifacts, the way spatiotemporal patterns are extracted from a dataset, the reliance on the wrong method, or possibly dynamic and causal intermittency, may alter a DAG and explain the

occurrence of spurious links. These topics could be studied more in-depth, based on virtual hydrological experiments. With multivariate methods, a spurious link may have negative feedback. The time-dependencies of the spurious driver are removed from the response variable while it should not. From this fact, combined with the idea the causes of spurious links are hardly identifiable or manageable in real case studies, Chapter 6 justifies the constraint of possible causal connections based on what links are conceptually or physically acceptable (as Rinderer et al., 2018). Doing so is a way to account for Hume's contiguity criterion that the cause and its effects should coexist in a domain that allows interaction (d23). It is also in phase with the idea that Hume's mysterious necessary connection (d25), or the proper justification of our beliefs (section 2.3.2, d6), is an account of the mechanism (section 2.4.3.2). Lastly, it is consistent with the idea that causality in the broad sense integrates prior knowledge to improve its understanding and that causality emerges from the constraining potentialities to one single representation (section 2.5).

Still, there is a fine line between constraining a method because of its limitations for the sake of reliability and constraining empirical facts to our own causal perceptions of reality (section 2.4.4). Causal inference from time-series is potentially a powerful approach, but not a magic wand to reveal causal interactions. It remains a modeling approach requiring a modeling philosophy that should be proactively debated. This philosophy must go hand in hand with an ethic. Chapter 6 shows that, for any pair of variables, there is likely to be today a combination of methods and data manipulation that would lead to the conclusion of a significant association. Hence, p-values do not make causal claims. It is, therefore, the responsibility of the scientist to use several methods, hypotheses, parameters, and data arrangements that could make sense and to report the variability of the results, and to conclude, beyond empirical induction (d13), following the logic of abduction (d14), i.e., the inference to the best hypothesis. Generally, confidence in the results for specific causal links increases with the redundancies between the approaches (as for Chapter 5), their robustness, the fitness-of-purpose of the causal inference method for the study case, and the consistencies when confronting the results to the perceptual or scientific knowledge of the system. If substantial inconsistencies between empirical methods and perceptions remain, this is good news rather than disillusionment, as science is on the verge of progressing either by fixing its method or revisiting its perception.

7.5 Causality from Empirical Time-Series Analysis

Ultimately, what does it take to understand complex hydrological systems based on empirical time-series analysis? First, the thesis proposal advanced the idea that a nonlinear causal inference method was needed to unravel the causality of a hydrological system (section 1.2.1). The reflections and experimentation during the Ph.D. quickly shifted to the idea that there was a need to go further and broadened the question by proposing a philosophical framework (Chapter 2). Two case studies based on empirical time series analysis were proposed, although not directly based on causal inference methods (Chapter 4 and 5). They nevertheless aim at a better causal understanding of complex hydrological systems. Finally, a detailed study of causal inference methods on hydrological data revealed their potential and limitations. From the rich experiences of the thesis, one may conclude that the following elements are necessary to understand complex hydrological systems from time-series analysis.

Data – It is undoubtedly the most obvious point since there is no empirical analysis without data. In terms of causality, the thesis shows the importance of having access to the set of explanatory variables in order to be able to infer the causality of a system. Ideally, a global data network should be built, combining data from remote sensing and field observatories, using standardized acquisition and documentation methods. Hydrologists have always suffered from the lack of observability of hydrological systems. While still far from ideal, technological advances and developing community impulses suggest that hydrology will move closer to this ideal in the years to come (Beven et al., 2020; Sivapalan and Blöschl, 2017).

Computing power – Alongside technological progress in data acquisition, computing power is essential. Nonlinear data-based methods that make the least assumptions about dynamics or probability distributions generally make use of nonparametric testing based on bootstrapping or data shuffling (Chapter 4, Chapter 6). These approaches required a high computation and parallelization, even on relatively small sets of variables and daily resolution. Since it is argued that the sensitivity of the causal inference method should be tested, the use of high-performance computing very quickly becomes a necessity for those who engage in causal inference with these methods.

Willingness to understand - Today, empirical methods are still often regarded as methods that precisely allow us to abstain from the causal understanding of a system. These empirical methods aim at the most precise possible operational predictions. The term "cause" is no longer used, but rather "predictor". However, in a changing system, these operational models can quickly become obsolete if they are not based on causal variables and some understanding of the mechanism. This thesis demonstrates that approaching causality from an empirical perspective is feasible in many ways, but what is needed first and foremost is a willingness to understand. It involves studying empirical methods and questioning how they fit into the philosophy of empirical causality. It involves studying the data, manipulating them, and asking the question cause or predictor?

Clear purpose – Solving causality to solve causality can quickly become a slippery slope for an analyst who forgets where he or she wants to sail to. Causality is certainly easier to resolve when we remember why we want to resolve it, based on an established and precise research question. The objectives bring focus, relevant concepts, and scope to evaluate the results. For instance, Chapters 5 and 6 test methods and raise awareness of their limitations and potential to the hydrogeological community. They lack finality, as the thesis aims to resolve causality, so to speak, which leads to plural results and a sense of pessimism. However, if the objective was, for example, to predict drip discharge patterns in a causally consistent manner, many combinations of clustering models and causal models could have been excluded if an objective function had constrained the whole process.

Getting to know and organizing the data – No empirical method is a magical device that eats data to spit out causality. Certainly less than physically-based models, causal inference methods nevertheless make assumptions, mainly about dynamics, e.g., whether it is stationary or not, intermittent or continuous, linear or not, deterministic or stochastic. Testing these hypotheses on univariate time series is an excellent way to become familiar with the data (e.g., Chapter 4), as well as decomposing and visualizing data. Exploratory data analysis allows one to consider how to pre-process the data before causal inference. It aims at making choices and solving the questions of the appropriate temporal or spectral domain for the analysis, knowing which data should or should not be considered, removing redundancies to get rid of excessive dimensionality, and extracting

spatiotemporal patterns that are relevant and as exhaustive as possible in the data set (e.g., Chapter 5).

Causal inference methods – Exploratory data analysis and purpose may orient to the choice of a causal inference method. Nevertheless, several methods can be used, each for a legitimate reason, the last being that we may not know which one is the most suitable. This thesis investigated methods that seek short-term lagged dependencies between variables. Other methods in the literature attempt to resolve contemporaneous dependencies or to infer causality from the spectral domain. Given that hydrological processes coexist on various time scales, the latter is also of particular interest. To some extent, many empirical models can be suited for causal inference since it is about looking for predictability while removing or controlling the effect of potential shared drivers, but we expect them to remain simple.

Balancing constraints and flexibility – It has been argued that causality emerges from practical, logical, and social constraints, but not from dogmatism. Causality also arises from the convergence of different approaches and perspectives towards the same generalization. We expect that this generalization remains a robust and tool for various purposes. In general, all chapters valued varying the perspective, the methods, or their parameters. Similarly, in an empirical study case, it goes without saying that if clear-cut choices, even if motivated, are made along the way, the analysis will logically lead to a single result. The empirical analysis's sensitivity needs to be tested, especially since the thesis has shown that they are potentially sensitive. It is up to the empirical modeler to define alternatives throughout the entire workflow, be it in the objective, its evaluation function, the definition of concepts, the way data is organized, the methods of causal inference, their parameters, or the user's constraints on potential causal links. This analysis must be judged on the extent of what is possible and reasonable because the robustness of an approach must not be validated or refuted based on meaningless alternatives. It follows the same logic of uncertainty and sensitivity analysis of physically-based models. However, with empirical approaches, this is quite an art because the choices to be made with an empirical approach are often case-specific, thus, empirical, and not easily related to the environment and its physical mechanism. It is possible to find guidelines for empirical methods in the literature, but these have not been thought out or suggested for a hydrological problem. It takes time, and often a community, to build expertise around new empirical methods in their

scientific domain. Virtual experiments may help develop guidance, with the risk that guidance may not apply in real cases (Chapter 6).

Building a case for science – Whatsoever comes out in empirical causal analysis is not causality as a whole (Chapter 2). At best, it is an organized representation of a system (Chapter 4, 5, 6), which, in the case of the causal inference method (Chapter 6), reports the system organization as the structure of causal associations between variables. For an association to become causality, it will be necessary to cross-reference points of view and build a case for the exposed causal relationships. While varying empirical approaches and making them more robust may help, this cannot ultimately be done by empirical time series analysis alone. At some point, we will always expect a causal claim a comment on the mechanism or consistency with physically-based models. One will also expect a claim to be verified with field experiments to the extent that it is possible. Whatever the means of adding evidence to the case, the empirical modelers may not have it in their skill set. Also, the strength of empirical modelers is the broad applicability of their knowledge and skills. However, with new study cases and different systems, their expertise and knowledge of the system behavior diminish, as the familiarity with the scientific literature or data type. This is even more true as soon as they move to another scientific field or research question. For all these reasons, the empirical modeler benefits from networking and allying with researchers that are willing to enter in a win-win relationship trading skills, knowledge, and vantage points, and perhaps, more fruitfully from the very beginning of the causal inference process that had just been described. More than a group of people, causality is a community challenge. All these elements mentioned above in this roadmap guide a modeler. These empirical discovery tips are generalizable for a community. Since causality is about extrapolable generalization, this thesis has inferred a consistent frame of Causality from empirical time-series analysis while inferring the hydrological system's causality from time-series is under progress.

References

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., Hannah, et al.: Human domination of the global water cycle absent from depictions and perceptions, Nat. Geosci., 12(7), 533–540, doi:10.1038/s41561-019-0374-y, 2019.

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., et al.: Machine learning for neuroimaging with scikit-learn, Front. Neuroinformatics, 8, doi:10.3389/fninf.2014.00014, 2014.

Aghabozorgi, S., Seyed Shirkhorshidi, A. and Ying Wah, T.: Time-series clustering – A decade review, Inf. Syst., 53, 16–38, doi:10.1016/j.is.2015.04.007, 2015.

Akaike, H.: A new look at the statistical model identification, IEEE Trans. Autom. Control, 19(6), 716–723, doi:10.1109/TAC.1974.1100705, 1974.

Aksoy, H. and Wittenberg, H.: Nonlinear baseflow recession analysis in watersheds with intermittent streamflow, Hydrol. Sci. J., 56(2), 226–237, doi:10.1080/02626667.2011.553614, 2011.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration: guidelines for computing crop water requirements, edited by Food and Agriculture Organization of the United Nations, Food and Agriculture Organization of the United Nations, Rome., 1998.

Andréassian, V., Perrin, C., Parent, E. and Bárdossy, A.: The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?, Hydrol. Sci. J., 55(6), 849–856, doi:10.1080/02626667.2010.506050, 2010.

Angelini, P.: Correlation and spectral analysis of two hydrogeological systems in Central Italy, Hydrol. Sci. J., 42(3), 425–438, doi:10.1080/02626669709492038, 1997.

Angioni, L.: Aristotle's Definition of Scientific Knowledge (APo 71b 9–12), Hist. Philos. Log. Anal., 19(1), 140–166, doi:10.30965/26664275-01901010, 2016.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M. and Perona, I.: An extensive comparative study of cluster validity indices, Pattern Recognit., 46(1), 243–256, doi:10.1016/j.patcog.2012.07.021, 2013.

Aristotle: The complete works of Aristotle: the revised Oxford translation. Volume One, edited by J. Barnes, Princeton Univ. Press, Princeton, NJ., 1995.

Arthur, D. and Vassilvitskii, S.: K-means++: The Advantages of Careful Seeding, in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. [online] Available from: http://dl.acm.org/citation.cfm?id=1283383.1283494, 2007.

Audebert, M., Clément, R., Touze-Foltz, N., Günther, T., Moreau, S. and Duquennoi, C.: Time-lapse ERT interpretation methodology for leachate injection monitoring based on multiple inversions and a clustering strategy (MICS), J. Appl. Geophys., 111, 320–333, doi:10.1016/j.jappgeo.2014.09.024, 2014.

Bailly-Comte, V., Jourde, H., Roesch, A., Pistre, S. and Batiot-Guilhe, C.: Time series analyses for Karst/River interactions assessment: Case of the Coulazou river (southern France), J. Hydrol., 349(1), 98–114, doi:10.1016/j.jhydrol.2007.10.028, 2008.

Bakalowicz, M.: Karst groundwater: a challenge for new resources, Hydrogeol. J., 13(1), 148–160, doi:10.1007/s10040-004-0402-9, 2005.

Baker, A.: Simplicity, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2016/entries/simplicity/ (Accessed 3 December 2019), 2016.

Baker, V. R.: Debates—Hypothesis testing in hydrology: Pursuing certainty versus pursuing uberty, Water Resour. Res., 53(3), 1770–1778, doi:10.1002/2016WR020078, 2017.

Baldassarre, G. D., Viglione, A., Carr, G., Kuil, L., Salinas, J. L. and Blöschl, G.: Socio-hydrology: conceptualising human-flood interactions, Hydrol. Earth Syst. Sci., 17(8), 3295–3303, doi:https://doi.org/10.5194/hess-17-3295-2013, 2013.

Banton, O., Cimon, M.-A. and Seguin, M.-K.: Mapping Field-Scale Physical Properties of Soil with Electrical Resistivity, Soil Sci. Soc. Am. J., 61(4), 1010–1017, doi:10.2136/sssaj1997.03615995006100040003x, 1997.

Barker, R. and Moore, J.: The application of time-lapse electrical tomography in groundwater studies, Lead. Edge, 17(10), 1454–1458, doi:10.1190/1.1437878, 1998.

Berger, P. L., Luckmann, T. and Zifonun, D.: The social construction of reality, 1967.

Bergson, H.: Creative Evolution., 1907.

Berkeley, G.: A treatise concerning the principles of human knowledge., 1710.

Berkhin, P.: A Survey of Clustering Data Mining Techniques, in Grouping Multidimensional Data, edited by J. Kogan, C. Nicholas, and M. Teboulle, pp. 25–71, Springer-Verlag, Berlin/Heidelberg., 2006.

Berne, A., Uijlenhoet, R. and Troch, P. A.: Similarity analysis of subsurface flow response of hillslopes with complex geometry, Water Resour. Res., 41(9), doi:10.1029/2004WR003629, 2005.

Berry, C. J.: The Kantian Revolution, in Hume, Hegel and Human Nature, edited by C. J. Berry, pp. 43–53, Springer Netherlands, Dordrecht., 1982.

von Bertalanffy, L.: General system theory: foundations, development, applications, G. Braziller. [online] Available from: https://books.google.be/books?id=mWZQAAAAMAAJ, 1968.

Beven, K.: Changing ideas in hydrology — The case of physically-based models, J. Hydrol., 105(1), 157–172, doi:10.1016/0022-1694(89)90101-7, 1989.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Adv. Water Resour., 16(1), 41–51, doi:10.1016/0309-1708(93)90028-E, 1993.

Beven, K.: Uniqueness of place and process representations in hydrological modelling, Hydrol. Earth Syst. Sci., 4(2), 203–213, doi:https://doi.org/10.5194/hess-4-203-2000, 2000.

Beven, K.: Towards a coherent philosophy for modelling the environment, Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci., 458(2026), 2465–2484, doi:10.1098/rspa.2002.0986, 2002a.

Beven, K.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, Hydrol. Process., 16(2), 189–206, doi:10.1002/hyp.343, 2002b.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320(1), 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006a.

Beven, K.: Searching for the Holy Grail of scientific hydrology: $Q_t=(S, R, \Delta t)A$ as closure, Hydrol. Earth Syst. Sci., 10(5), 609–618, doi:https://doi.org/10.5194/hess-10-609-2006, 2006b.

Beven, K.: Causal models as multiple working hypotheses about environmental processes, Comptes Rendus Geosci., 344(2), 77–88, doi:10.1016/j.crte.2012.01.005, 2012a.

Beven, K. : Rainfall-runoff modelling: the primer, 2nd ed., Wiley-Blackwell, Chichester, West Sussex ; Hoboken, NJ., 2012b.

Beven, K.: Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, Hydrol. Sci. J., 61(9), 1652–1665, doi:10.1080/02626667.2015.1031761, 2016.

Beven, K.: On hypothesis testing in hydrology: Why falsification of models is still a really good idea, Wiley Interdiscip. Rev. Water, 5(3), e1278, doi:10.1002/wat2.1278, 2018.

Beven, K.: How to make advances in hydrological modelling, Hydrol. Res., 50(6), 1481–1494, doi:10.2166/nh.2019.134, 2019a.

Beven, K.: Towards a methodology for testing models as hypotheses in the inexact sciences, Proc. R. Soc. Math. Phys. Eng. Sci., 475(2224), 20180862, doi:10.1098/rspa.2018.0862, 2019b.

Beven, K. and Cloke, H. L.: Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al., Water Resour. Res., 48(1), doi:10.1029/2011WR010982, 2012.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49(6), 3071–3092, doi:10.1002/wrcr.20156, 2013.

Beven, K., Smith, P. J. and Freer, J. E.: So just why would a modeller choose to be incoherent?, J. Hydrol., 354(1), 15–32, doi:10.1016/j.jhydrol.2008.02.007, 2008.

Beven, K., Smith, P., Westerberg, I. and Freer, J.: Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al., Water Resour. Res., 48(11), doi:10.1029/2012WR012282, 2012.

Beven, K., Asadullah, A., Bates, P., Blyth, E., Chappell, N., Child, S., Cloke, H., et al.: Developing observational methods to drive future hydrological science: Can we make a start as a community?, Hydrol. Process., 34(3), 868–873, doi:10.1002/hyp.13622, 2020.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., Roo, A. de, et al.: Hyper-resolution global hydrological modelling: what is next?, Hydrol. Process., 29(2), 310–320, doi:10.1002/hyp.10391, 2015.

Biswal, B. and Marani, M.: Geomorphological origin of recession curves, Geophys. Res. Lett., 37(24), doi:10.1029/2010GL045415, 2010.

Blöschl, G.: Hydrologic synthesis: Across processes, places, and scales, Water Resour. Res., 42(3), doi:10.1029/2005WR004319, 2006.

Blöschl, G., Ed.: Runoff prediction in ungauged basins: synthesis across processes, places and scales, Cambridge University Press, Cambridge., 2013.

Blöschl, G.: Debates—Hypothesis testing in hydrology: Introduction, Water Resour. Res., 53(3), 1767–1769, doi:10.1002/2017WR020584, 2017.

Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, Hydrol. Process., 9(3–4), 251–290, doi:10.1002/hyp.3360090305, 1995.

Blöschl, G. and Zehe, E.: On hydrological predictability, Hydrol. Process., 19(19), 3923–3929, doi:10.1002/hyp.6075, 2005.

Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T. and Viglione, A.: Runoff prediction in ungauged basins: synthesis across processes, places and scales, Cambridge University Press., 2013.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., et al.: Twenty-three unsolved problems in hydrology (UPH) – a community perspective, Hydrol. Sci. J., 64(10), 1141–1158, doi:10.1080/02626667.2019.1620507, 2019.

Blume, T., Zehe, E. and Bronstert, A.: Rainfall—runoff response, event-based runoff coefficients and hydrograph separation, Hydrol. Sci. J., 52(5), 843–862, doi:10.1623/hysj.52.5.843, 2007.

Bogaard, T. A., Illingworth, S. M., Stewart, I. and van Manen, S., Eds.: Special issue | Effective Science Communication and Education in Hydrology and Natural Hazards (NHESS/HESS inter-journal SI), [online] Available from: https://www.hydrol-earth-syst-sci.net/special_issue7_639.html (Accessed 12 March 2020), 2017.

Bogaart, P. W., Velde, Y. van der, Lyon, S. W. and Dekker, S. C.: Streamflow recession patterns can help unravel the role of climate and humans in landscape co-evolution, Hydrol. Earth Syst. Sci., 20(4), 1413–1432, doi:https://doi.org/10.5194/hess-20-1413-2016, 2016.

Boholm, A.: Comparative studies of risk perception: a review of twenty years of research, J. Risk Res., 1(2), 135–163, doi:10.1080/136698798377231, 1998.

Boholm, M.: Risk and Causality in Newspaper Reporting, Risk Anal., 29(11), 1566–1577, doi:10.1111/j.1539-6924.2009.01296.x, 2009.

Bonniver, I., Reck, S. and Hallet, V.: Rochefort-Nassogne 59/3-4, Notice explicative, [online] Available from: http://environnement.wallonie.be/cartosig/cartehydrogeo/document/Notice_5934.pdf, 2013.

Boussinesq, J.: Sur un mode simple d'écoulement des nappes d'eau d'infiltration à lit horizontal, avec rebord vertical tout autour lorsqu'une partie de ce rebord est enlevée depuis la surface jusqu'au fond., 1903.

Box, G. E. P.: Science and Statistics, J. Am. Stat. Assoc., 71(356), 791–799, doi:10.1080/01621459.1976.10480949, 1976.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C.: Time series analysis: forecasting and control, J. Wiley & Sons, Hoboken, N.J. [online] Available from: http://www.123library.org/book_details/?id=30721 (Accessed 10 February 2020), 2008.

Boyaci Gülenç, N. P.: An Enquiry on Physis–Nomos Debate: Sophists, Synth. Philos., 31(1), doi:10.21464/sp31103, 2016.

Bracken, L. J., Wainwright, J., Ali, G. A., Tetzlaff, D., Smith, M. W., Reaney, S. M. and Roy, A. G.: Concepts of hydrological connectivity: Research approaches, pathways and future agendas, Earth-Sci. Rev., 119, 17–34, doi:10.1016/j.earscirev.2013.02.001, 2013.

Bradford, R. A., O'Sullivan, J. J., Craats, I. M. van der, Krywkow, J., Rotko, P., Aaltonen, J., Bonaiuto, et al.: Risk perception – issues for flood management in Europe, Nat. Hazards Earth Syst. Sci., 12(7), 2299–2309, doi:https://doi.org/10.5194/nhess-12-2299-2012, 2012.

Bras, R. L.: A Brief History of Hydrology: The Robert E. Horton Lecture, Bull. Am. Meteorol. Soc., 80(6), 1151–1164, doi:10.1175/1520-0477-80.6.1151, 1999.

Bras, R.L. Complexity and organization in hydrology: A personal view. Water Resources Research, 51, 6532–6548, doi:10.1002/2015WR016958, 2015.

Brewster, H.: The river gods of Greece: myths and mountain waters in the Hellenic world, IB Tauris., 1997.

Brodie, R. and Hostetler, S.: A review of techniques for analysing baseflow from stream hydrographs, in Proceedings of the NZHS-IAH-NZSSS 2005 conference, vol. 28, Auckland New Zealand., 2005.

Bronowski, J.: Science and Human Values, Julian Messner Inc., New York., 1956.

Brown, R. and Fish, D.: The psychological causality implicit in language, Cognition, 14(3), 237–273, 1983.

Brulein, J.: Approche combinée d'inférence causale et d'apprentissage machine pour la modélisation de variables hydro-géophysiques du sytème karstique de la Lhomme (Rochefort, Belgique)., Master thesis, UCL - Université Catholique de Louvain, Louvain-la-Neuve., 2019.

Brutsaert, W.: Long-term groundwater storage trends estimated from streamflow records: Climatic perspective: LONG-TERM GROUNDWATER STORAGE TRENDS, Water Resour. Res., 44(2), doi:10.1029/2007WR006518, 2008.

Brutsaert, W. and Nieber, J. L.: Regionalized drought flow hydrographs from a mature glaciated plateau, Water Resour. Res., 13(3), 637–643, doi:10.1029/WR013i003p00637, 1977.

Buchecker, M., Salvini, G., Baldassarre, G. D., Semenzin, E., Maidl, E. and Marcomini, A.: The role of risk perception in making flood risk management more effective, Nat. Hazards Earth Syst. Sci., 13(11), 3013–3030, doi:https://doi.org/10.5194/nhess-13-3013-2013, 2013.

Burn, D. H.: Perceptions of flood risk: A case study of the Red River Flood of 1997, Water Resour. Res., 35(11), 3451–3458, doi:10.1029/1999WR900215, 1999.

Burt, T. P. and McDonnell, J. J.: Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses, Water Resour. Res., 51(8), 5919–5928, doi:10.1002/2014WR016839, 2015.

Camelbeeck, T., van Ruymbeke, M., Quinif, Y., Vandycke, S., de Kerchove, E. and Ping, Z.: Observation and interpretation of fault activity in the Rochefort cave (Belgium), Tectonophysics, 581, 48–61, doi:10.1016/j.tecto.2011.09.027, 2012.

Caston, V.: Intentionality in Ancient Philosophy, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2019/entries/intentionality-ancient/ (Accessed 25 February 2020), 2019.

Caterina, D., Beaujean, J., Robert, T. and Nguyen, F.: A comparison study of different image appraisal tools for electrical resistivity tomography, Surf. Geophys., 11(6), 639–657, doi:10.3997/1873-0604.2013022, 2013.

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., et al.: Virtual laboratories: new opportunities for collaborative water science, Hydrol. Earth Syst. Sci., 19(4), 2101–2117, doi:https://doi.org/10.5194/hess-19-2101-2015, 2015.

Chambers, J. E., Wilkinson, P. B., Wardrop, D., Hameed, A., Hill, I., Jeffrey, C., Loke, M. H., et al.: Bedrock detection beneath river terrace deposits using three-dimensional electrical resistivity tomography, Geomorphology, 177–178, 17–25, doi:10.1016/j.geomorph.2012.03.034, 2012.

Chambers, J. E., Wilkinson, P. B., Penn, S., Meldrum, P. I., Kuras, O., Loke, M. H. and Gunn, D. A.: River terrace sand and gravel deposit reserve estimation using three-dimensional electrical resistivity tomography for bedrock surface detection, J. Appl. Geophys., 93, 25–32, doi:10.1016/j.jappgeo.2013.03.002, 2013.

Chambers, J. E., Meldrum, P. I., Wilkinson, P. B., Ward, W., Jackson, C., Matthews, B., Joel, P., et al.: Spatial monitoring of groundwater drawdown and rebound associated with quarry

dewatering using automated time-lapse electrical resistivity tomography and distribution guided clustering, Eng. Geol., 193, 412–420, doi:10.1016/j.enggeo.2015.05.015, 2015.

Chapman, T.: A comparison of algorithms for stream flow recession and baseflow separation, Hydrol. Process., 13(5), 701–714, 1999.

Chappell, S.-G.: Plato on Knowledge in the *Theaetetus*, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2019/entries/plato-theaetetus/ (Accessed 13 December 2019), 2019.

Chen, B. and Krajewski, W. F.: Recession analysis across scales: The impact of both random and nonrandom spatial variability on aggregated hydrologic response, J. Hydrol., 523, 97–106, doi:10.1016/j.jhydrol.2015.01.049, 2015.

Chen, Z., Auler, A. S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., Jiang, G., et al.: The World Karst Aquifer Mapping project: concept, mapping procedure and map of Europe, Hydrogeol. J., 25(3), 771–785, doi:10.1007/s10040-016-1519-3, 2017.

Chong, D. and Druckman, J. N.: Framing Theory, Annu. Rev. Polit. Sci., 10(1), 103–126, doi:10.1146/annurev.polisci.10.072805.103054, 2007.

Chow, V. T., Maidment, D. R. and Mays, L. W.: Applied Hydrology, McGraw-Hill. [online] Available from: https://books.google.be/books?id=rY_yMgEACAAJ, 1988.

Christofides, A. and Koutsoyiannis, D.: Causality in climate and hydrology, Eur. Geosci. Union Gen. Assem. 2011 Geophys. Res. Abstr., 13, 2011.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44(12), W00B02, doi:10.1029/2007WR006735, 2008.

Clark, M. P., Rupp, D. E., Woods, R. A., Meerveld, H. J. T., Peters, N. E. and Freer, J. E.: Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale, Hydrol. Process., 23(2), 311–319, doi:10.1002/hyp.7154, 2009.

Clark, M. P., Kavetski, D. and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47(9), doi:10.1029/2010WR009827, 2011.

Clark, M. P., Kavetski, D. and Fenicia, F.: Reply to comment by K. Beven et al. on "Pursuing the method of multiple working hypotheses for hydrological modeling," Water Resour. Res., 48(11), doi:10.1029/2012WR012547, 2012.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51(4), 2498–2514, doi:10.1002/2015WR017198, 2015.

Clifford, N. J.: Hydrology: the changing paradigm, Prog. Phys. Geogr. Earth Environ., 26(2), 290–301, doi:10.1191/0309133302pp337pr, 2002.

Cohen, S. M.: Aristotle's Metaphysics, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2016/entries/aristotle-metaphysics/ (Accessed 11 February 2020), 2016.

Creutzfeldt, B., Troch, P. A., Güntner, A., Ferré, T. P. A., Graeff, T. and Merz, B.: Storage-discharge relationships at different catchment scales based on local high-precision gravimetry, Hydrol. Process., 28(3), 1465–1475, doi:10.1002/hyp.9689, 2014.

Cukier, R. I., Levine, H. B. and Shuler, K. E.: Nonlinear sensitivity analysis of multiparameter model systems, J. Comput. Phys., 26(1), 1–42, doi:10.1016/0021-9991(78)90097-9, 1978.

Curd, P.: Presocratic Philosophy, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/sum2019/entries/presocratics/, 2019.

De Pierris, G. and Friedman, M.: Kant and Hume on Causality, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2018/entries/kant-hume-causality/ (Accessed 12 November 2019), 2018.

Delacy, P. H.: The Problem of Causation in Plato's Philosophy, Class. Philol., 34(2), 97–115, 1939.

Delforge, D., Vanclooster, M., Van Camp, M., Poulain, A., Watlet, A., Hallet, V., Kaufmann, O. and Francis, O.: Retrieving hydrological connectivity from empirical causality in karst systems. [online] Available from: https://dial.uclouvain.be/pr/boreal/object/boreal:184256 (Accessed 29 August 2017), 2017.

Delforge, D., Vanclooster, M. and Van Camp, M.: Hydrograph separation using recursive filters and causal objective function, in Geophysical Research Abstracts, vol. 20. [online] Available from: https://dial.uclouvain.be/pr/boreal/object/boreal:196934 (Accessed 8 April 2020), 2018.

Delforge, D., Vanclooster, M., Camp, M. V., Hallet, V., Kaufmann, O., Poulain, A. and Watlet, A.: Hydrological connectivity From Causal analysis of time series in the Lhomme Karst System (Belgium), [online] Available from: https://figshare.com/articles/Hydrological_connectivity_From_Causal_analysis_of_time_series_in_the_Lhomme_Karst_System_Belgium_/8025167, 2019.

Delforge, D., Muñoz-Carpena, R., Van Camp, M. and Vanclooster, M.: A Parsimonious Empirical Approach to Streamflow Recession Analysis and Forecasting, Water Resour. Res., 56(2), e2019WR025771, doi:10.1029/2019WR025771, 2020.

Delobbe, L., Watlet, A., Wilfert, S. and Camp, M. V.: Exploring the use of underground gravity monitoring to evaluate radar estimates of heavy rainfall, Hydrol. Earth Syst. Sci., 23(1), 93–105, doi:https://doi.org/10.5194/hess-23-93-2019, 2019.

Dempster, A. P., Laird, N. M. and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, J. R. Stat. Soc. Ser. B Methodol., 39(1), 1–38, 1977.

Descartes, R. and Renault, L.: Discours de la méthode, Flammarion, Paris., 2016.

Dewandel, B., Lachassagne, P., Bakalowicz, M., Weng, P. and Al-Malki, A.: Evaluation of aquifer thickness by analysing recession hydrographs. Application to the Oman ophiolite hard-rock aquifer, J. Hydrol., 274(1), 248–269, doi:10.1016/S0022-1694(02)00418-3, 2003.

Deyle, E. R. and Sugihara, G.: Generalized Theorems for Nonlinear State Space Reconstruction, PLoS ONE, 6(3), doi:10.1371/journal.pone.0018295, 2011.

Di Giuseppe, M. G., Troiano, A., Troise, C. and De Natale, G.: k-Means clustering as tool for multivariate geophysical data analysis. An application to shallow fault zone imaging, J. Appl. Geophys., 101, 108–115, doi:10.1016/j.jappgeo.2013.12.004, 2014.

Di Giuseppe, M. G., Troiano, A., Patella, D., Piochi, M. and Carlino, S.: A geophysical k-means cluster analysis of the Solfatara-Pisciarelli volcano-geothermal system, Campi Flegrei (Naples, Italy), J. Appl. Geophys., 156, 44–54, doi:10.1016/j.jappgeo.2017.06.001, 2018.

Doetsch, J., Linde, N., Coscia, I., Greenhalgh, S. A. and Green, A. G.: Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data, GEOPHYSICS, 75(6), G53–G64, doi:10.1190/1.3496476, 2010.

Dooge, J.: The hydrologic cycle as a closed system, Int. Assoc. Sci. Hydrol. Bull., 13(1), 58–68, doi:10.1080/02626666809493568, 1968.

Dooge, J.: Linear Theory of Hydrologic Systems, Agricultural Research Service, U.S. Department of Agriculture., 1973.

Dooge, J.: General report on model structure and classification, Logist. Benefits Using Math. Models Hydrol. Water Resour. Syst., 1–21, 1978.

Dooge, J..: Looking for hydrologic laws, Water Resour. Res., 22(9S), 46S-58S, doi:10.1029/WR022i09Sp0046S, 1986.

Dooge, J.: Searching for Simplicity in Hydrology, Surv. Geophys., 18(5), 511–534, doi:10.1023/A:1006557801884, 1997.

Dralle, D. N., Karst, N. J., Charalampous, K., Veenstra, A. and Thompson, S. E.: Event-scale power law recession analysis: quantifying methodological uncertainty, Hydrol. Earth Syst. Sci., 21(1), 65–81, doi:https://doi.org/10.5194/hess-21-65-2017, 2017.

Dubois, C., Quinif, Y., Baele, J.-M., Barriquand, L., Bini, A., Bruxelles, L., Dandurand, G., Havron, C., Kaufmann, O., Lans, B. and others: The process of ghost-rock karstification and its role in the formation of cave systems, Earth-Sci. Rev., 131, 116–148, 2014.

Dunne, T. and Black, R. D.: An Experimental Investigation of Runoff Production in Permeable Soils, Water Resour. Res., 6(2), 478–490, doi:10.1029/WR006i002p00478, 1970.

Eagleson, P. S.: Dynamic Hydrology, McGraw-Hill. [online] Available from: https://books.google.be/books?id=3rcPAQAAIAAJ, 1970.

Eagleson, P. S.: Ecohydrology: Darwinian Expression of Vegetation Form and Function, Camb. Core, doi:10.1017/CBO9780511535680, 2002.

Eckhardt, K.: « How to Construct Recursive Digital Filters for Baseflow Separation ». Hydrological Processes 19(2): 507-15. https://doi.org/10.1002/hyp.5675, 2005.

Ehret, U., Gupta, H. V., Sivapalan, M., Weijs, S. V., Schymanski, S. J., Blöschl, G., Gelfan, A. N., et al.: Advancing catchment hydrology to deal with predictions under change, Hydrol. Earth Syst. Sci., 18(2), 649–671, doi:https://doi.org/10.5194/hess-18-649-2014, 2014.

Elwaseif, M. and Slater, L.: Quantifying tomb geometries in resistivity images using watershed algorithms, J. Archaeol. Sci., 37(7), 1424–1436, doi:10.1016/j.jas.2010.01.002, 2010.

Elwaseif, M. and Slater, L.: Improved Resistivity Imaging of Targets with Sharp Boundaries Using an Iterative Disconnect Procedure, J. Environ. Eng. Geophys., 17(2), 89–101, doi:10.2113/JEEG17.2.89, 2012.

Falcon, A.: Aristotle on Causality, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/spr2019/entries/aristotle-causality/ (Accessed 6 November 2019), 2019.

Falkenmark, M.: The Greatest Water Problem: The Inability to Link Environmental Security, Water Security and Food Security, Int. J. Water Resour. Dev., 17(4), 539–554, doi:10.1080/07900620120094073, 2001.

Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: Is the groundwater reservoir linear? Learning from data in hydrological modelling, Hydrol Earth Syst Sci, 10(1), 139–150, doi:10.5194/hess-10-139-2006, 2006.

Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(11), W11510, doi:10.1029/2010WR010174, 2011.

Fiandaca, G., Doetsch, J., Vignoli, G. and Auken, E.: Generalized focusing of time-lapse changes with applications to direct current and time-domain induced polarization inversions, Geophys. J. Int., 203(2), 1101–1112, doi:10.1093/gji/ggv350, 2015.

Flügel, W.-A.: Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany, Hydrol. Process., 9(3–4), 423–436, doi:10.1002/hyp.3360090313, 1995.

Ford, D. and Williams, P. W.: Karst hydrogeology and geomorphology, Rev. ed., John Wiley & Sons, Chichester, England ; a Hoboken, NJ., 2007.

Freeze, R. A. and Harlan, R. L.: Blueprint for a physically-based, digitally-simulated hydrologic response model, J. Hydrol., 9(3), 237–258, doi:10.1016/0022-1694(69)90020-1, 1969.

Frenzel, S. and Pompe, B.: Partial mutual information for coupling analysis of multivariate time series, Phys. Rev. Lett., 99(20), 204101, doi:10.1103/PhysRevLett.99.204101, 2007.

Fuchs, S., Karagiorgos, K., Kitikidou, K., Maris, F., Paparrizos, S. and Thaler, T.: Flood risk perception and adaptation capacity: a contribution to the socio-hydrology debate, Hydrol. Earth Syst. Sci., 21(6), 3183–3198, doi:https://doi.org/10.5194/hess-21-3183-2017, 2017.

Gallego, G., Cuevas, C., Mohedano, R. and Garcia, N.: On the Mahalanobis Distance Classification Criterion for Multidimensional Normal Distributions, IEEE Trans. Signal Process., 61(17), 4387–4396, doi:10.1109/TSP.2013.2269047, 2013.

Galton, F.: Natural inheritance., 1889.

Genelle, F., Sirieix, C., Riss, J. and Naudet, V.: Monitoring landfill cover by electrical resistivity tomography on an experimental site, Eng. Geol., 145–146, 18–29, doi:10.1016/j.enggeo.2012.06.002, 2012.

Goldscheider, N. and Drew, D., Eds.: Methods in karst hydrogeology, Taylor & Francis, Leiden ; New York., 2007.

Goodkind, J. M.: The superconducting gravimeter, Rev. Sci. Instrum., 70(11), 4131–4152, doi:10.1063/1.1150092, 1999.

Goodwell, A. E., Jiang, P., Ruddell, B. L. and Kumar, P.: Debates—Does Information Theory Provide a New Paradigm for Earth Science? Causality, Interaction, and Feedback, Water Resour. Res., 56(2), doi:10.1029/2019WR024940, 2020.

Got, J.-B.: Soil piping : detection, hydrological functioning and modeling. A case study in loess-derived soils in Belgium, UCL - Université Catholique de Louvain. [online] Available from: https://dial.uclouvain.be/pr/boreal/object/boreal:222920 (Accessed 27 July 2020), 2019.

Granger, C. W. J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods, Econometrica, 37(3), 424–438, doi:10.2307/1912791, 1969.

Grant, G. E. and Dietrich, W. E.: The frontier beneath our feet, Water Resour. Res., 53(4), 2605–2609, doi:10.1002/2017WR020835, 2017.

Grayson, R. B., Moore, I. D. and McMahon, T. A.: Physically based hydrologic modeling: 2. Is the concept realistic?, Water Resour. Res., 28(10), 2659–2666, doi:10.1029/92WR01259, 1992.

Green, C. H., Tunstall, S. M. and Fordham, M. H.: The Risks from Flooding: Which Risks and Whose Perception?, Disasters, 15(3), 227–236, doi:10.1111/j.1467-7717.1991.tb00456.x, 1991.

Green, W. H. and Ampt, G. A.: Studies on Soil Phyics., J. Agric. Sci., 4(1), 1–24, doi:10.1017/S0021859600001441, 1911.

Grene, M.: Aristotle and Modern Biology, in Topics in the Philosophy of Biology, vol. 27, pp. 3–36, Springer Netherlands, Dordrecht., 1976.

Günther, T., Rücker, C. and Spitzer, K.: Three-dimensional modelling and inversion of dc resistivity data incorporating topography - II. Inversion, Geophys. J. Int., 166(2), 506–517, doi:10.1111/j.1365-246X.2006.03011.x, 2006.

Gupta, H. V. and Razavi, S.: Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models, Water Resour. Res., 54(11), 8692–8717, doi:10.1029/2018WR022668, 2018.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G. and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48(8), doi:10.1029/2011WR011044, 2012.

Hall, F. R.: Base-Flow Recessions-A Review, Water Resour. Res., 4(5), 973–983, doi:10.1029/WR004i005p00973, 1968.

Hallet, V. and Meus, P.: Contexte hydrogéologique des systèmes karstiques de la région de Rochefort (Vallées de la Wamme et de la Lomme), in International symposium “Karst research challenges for the XXIst century” - Brussels, 30 September & Rochefort, 1st October 2011. The karst network of the Lomme river, Rochefort region : guide book, vol. 309, pp. 039–046, Institut royal des Sciences naturelles de Belgiques, Service géologique de Belgique, Rue Jenner 13, 1000 Bruxelles., 2011.

Harman, C. and Sivapalan, M.: A similarity framework to assess controls on shallow subsurface flow dynamics in hillslopes, Water Resour. Res., 45(1), doi:10.1029/2008WR007067, 2009a.

Harman, C. and Sivapalan, M.: Effects of hydraulic conductivity variability on hillslope-scale shallow subsurface flow response and storage-discharge relations, Water Resour. Res., 45(1), doi:10.1029/2008WR007228, 2009b.

Harman, C. and Troch, P. A.: What makes Darwinian hydrology “Darwinian”? Asking a different kind of question about landscapes, Hydrol. Earth Syst. Sci., 18(2), 417–433, doi:https://doi.org/10.5194/hess-18-417-2014, 2014.

Harman, C. J., Sivapalan, M. and Kumar, P.: Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics, Water Resour. Res., 45(9), doi:10.1029/2008WR007392, 2009.

Harte, J.: Toward a Synthesis of the Newtonian and Darwinian Worldviews, Phys. Today, 55(10), 29–34, doi:10.1063/1.1522164, 2002.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J. and Weiler, M.: Karst water resources in a changing world: Review of hydrological modeling approaches, Rev. Geophys., 52(3), 2013RG000443, doi:10.1002/2013RG000443, 2014.

Hasan, S., Troch, P. A., Boll, J. and Kroner, C.: Modeling the Hydrological Effect on Local Gravity at Moxa, Germany, J. Hydrometeorol., 7(3), 346–354, doi:10.1175/JHM488.1, 2006.

Heider, F.: The psychology of interpersonal relations, John Wiley & Sons Inc, Hoboken, NJ, US., 1958.

Hermans, T. and Irving, J.: Facies discrimination with electrical resistivity tomography using a probabilistic methodology: effect of sensitivity and regularisation, Surf. Geophys., 15(1), 13–25, doi:10.3997/1873-0604.2016047, 2017.

Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M. and Bhattacharya, J.: Causality detection based on information-theoretic approaches in time series analysis, Phys. Rep., 441(1), 1–46, doi:10.1016/j.physrep.2006.12.004, 2007.

Horton, R. E.: The Rôle of infiltration in the hydrologic cycle, Eos Trans. Am. Geophys. Union, 14(1), 446–460, doi:10.1029/TR014i001p00446, 1933.

Hrachowitz, M. and Clark, M. P.: HESS Opinions: The complementary merits of competing modelling philosophies in hydrology, Hydrol. Earth Syst. Sci., 21(8), 3953–3973, doi:https://doi.org/10.5194/hess-21-3953-2017, 2017.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., et al.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrol. Sci. J., 58(6), 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

Hsu, H.-L., Yanites, B. J., Chen, C. and Chen, Y.-G.: Bedrock detection using 2D electrical resistivity imaging along the Peikang River, central Taiwan, Geomorphology, 114(3), 406–414, doi:10.1016/j.geomorph.2009.08.004, 2010.

Huffaker, R., Muñoz-Carpena, R., Campo-Bescós, M. A. and Southworth, J.: Demonstrating correspondence between decision-support models and dynamics of real-world environmental systems, Environ. Model. Softw., 83, 74–87, doi:10.1016/j.envsoft.2016.04.024, 2016.

Huffaker, R., Bittelli, M. and Rosa, R.: Nonlinear Time Series Analysis with R, Oxford University Press., 2017.

Hume, D.: A Treatise of Human Nature, Oxford University Press., 1738.

Hume, D.: Philosophical Essays Concerning Human Understanding, A. Millar., 1748.

Hume, D.: Enquiries Concerning the Human Understanding and Concerning the Principles of Morals, Greenwood Press., 1777.

Ichikawa, J. J. and Steup, M.: The Analysis of Knowledge, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/ (Accessed 27 January 2020), 2018.

Infante, V., Gallardo, L. A., Montalvo-Arrieta, J. C. and Navarro de León, I.: Lithological classification assisted by the joint inversion of electrical and seismic data at a control site in northeast Mexico, J. Appl. Geophys., 70(2), 93–102, doi:10.1016/j.jappgeo.2009.11.003, 2010.

Ishola, K. S., Nawawi, M. N. M. and Abdullah, K.: Combining Multiple Electrode Arrays for Two-Dimensional Electrical Resistivity Imaging Using the Unsupervised Classification Technique, Pure Appl. Geophys., 172(6), 1615–1642, doi:10.1007/s00024-014-1007-4, 2015.

Islam, M. N. and Sivakumar, B.: Characterization and prediction of runoff dynamics: a nonlinear dynamical view, Adv. Water Resour., 25(2), 179–190, doi:10.1016/S0309-1708(01)00053-7, 2002.

Jachens, E. R., Rupp, D. E., Roques, C. and Selker, J. S.: Recession analysis revisited: impacts of climate on parameter estimation, Hydrol. Earth Syst. Sci., 24(3), 1159–1170, doi:https://doi.org/10.5194/hess-24-1159-2020, 2020.

Jacob, T., Bayer, R., Chery, J., Jourde, H., Moigne, N. L., Boy, J.-P., Hinderer, J., Luck, B. and Brunet, P.: Absolute gravity monitoring of water storage variation in a karst aquifer on the larzac plateau (Southern France), J. Hydrol., 359(1), 105–117, doi:10.1016/j.jhydrol.2008.06.020, 2008.

Kadić, A., Denić-Jukić, V. and Jukić, D.: Revealing hydrological relations of adjacent karst springs by partial correlation analysis, Hydrol. Res., 49(3), 616–633, doi:10.2166/nh.2017.064, 2018.

Kantz, H. and Schreiber, T.: Nonlinear Time Series Analysis, Camb. Core, doi:10.1017/CBO9780511755798, 2003.

Karlsen, R. H., Bishop, K., Grabs, T., Ottosson-Löfvenius, M., Laudon, H. and Seibert, J.: The role of landscape properties, storage and evapotranspiration on variability in streamflow recessions in a boreal catchment, J. Hydrol., 570, 315–328, doi:10.1016/j.jhydrol.2018.12.065, 2019.

Kaufmann, O., Bastin, C., Barcella, C., Watlet, A. and Van Ruymbeke, M.: Design and calibration of a system for monitoring highly variable dripwater flows in caves, GB2016, 129, 2016.

Kellens, W., Terpstra, T. and Maeyer, P. D.: Perception and Communication of Flood Risks: A Systematic Review of Empirical Research, Risk Anal., 33(1), 24–49, doi:10.1111/j.1539-6924.2012.01844.x, 2013.

Kelley, H. H.: The processes of causal attribution, Am. Psychol., 28(2), 107–128, doi:10.1037/h0034225, 1973.

Kember, G., Flower, A. C. and Holubeshen, J.: Forecasting river flow using nonlinear dynamics, Stoch. Hydrol. Hydraul., 7(3), 205–212, doi:10.1007/BF01585599, 1993.

Keogh, E. and Kasetty, S.: On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration, Data Min. Knowl. Discov., 7(4), 349–371, doi:10.1023/A:1024988512476, 2003.

Khatibi, R., Sivakumar, B., Ghorbani, M. A., Kisi, O., Koçak, K. and Farsadi Zadeh, D.: Investigating chaos in river stage and discharge time series, J. Hydrol., 414–415, 108–117, doi:10.1016/j.jhydrol.2011.10.026, 2012.

Kiraly, L.: Karstification and groundwater flow, Speleogenesis Evol. Karst Aquifers, 1(3), 155–192, 2003.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42(3), doi:10.1029/2005WR004362, 2006.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45(2), doi:10.1029/2008WR006912, 2009.

Kirchner, J. W.: Science, politics, and rationality in a partisan era, Water Resour. Res., 53(5), 3545–3549, doi:10.1002/2017WR020882, 2017.

Klemeš, V.: Empirical and causal models in hydrology, in Scientific Basis of Water-Resource Management, Washinton D.C. [online] Available from: http://www.itia.ntua.gr/en/docinfo/1075/ (Accessed 6 November 2019), 1982.

Klemeš, V.: Conceptualization and scale in hydrology, J. Hydrol., 65(1–3), 1–23, doi:10.1016/0022-1694(83)90208-1, 1983.

Klemeš, V.: Dilettantism in hydrology: Transition or destiny?, Water Resour. Res., 22(9S), 177S-188S, doi:10.1029/WR022i09Sp0177S, 1986a.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31(1), 13–24, doi:10.1080/02626668609491024, 1986b.

Koutsoyiannis, D.: Nonstationarity versus scaling in hydrology, J. Hydrol., 324(1), 239–254, doi:10.1016/j.jhydrol.2005.09.022, 2006a.

Koutsoyiannis, D.: On the quest for chaotic attractors in hydrological processes, Hydrol. Sci. J., 51(6), 1065–1091, doi:10.1623/hysj.51.6.1065, 2006b.

Koutsoyiannis, D.: HESS Opinions "A random walk on water," Hydrol. Earth Syst. Sci., 14(3), 585–601, doi:https://doi.org/10.5194/hess-14-585-2010, 2010.

Koutsoyiannis, D.: Hydrology and change, Hydrol. Sci. J., 58(6), 1177–1197, doi:10.1080/02626667.2013.804626, 2013.

Koutsoyiannis, D.: Reconciling hydrology with engineering, Hydrol. Res., 45(1), 2–22, doi:10.2166/nh.2013.092, 2014.

Koutsoyiannis, D.: Generic and parsimonious stochastic modelling for hydrology and beyond, Hydrol. Sci. J., 61(2), 225–244, doi:10.1080/02626667.2015.1016950, 2016.

Koutsoyiannis, D.: Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes, Hydrol. Sci. J. [online] Available from: https://www.tandfonline.com/doi/abs/10.1080/02626667.2019.1600700 (Accessed 24 February 2020), 2019.

Koutsoyiannis, D. and Montanari, A.: Negligent killing of scientific concepts: the stationarity case, Hydrol. Sci. J., 60(7–8), 1174–1183, doi:10.1080/02626667.2014.959959, 2015.

Koutsoyiannis, D., Mamassi, N. and Tegos, A.: Logical and illogical exegeses of hydrometeorological phenomena in ancient Greece, Water Sci. Technol. Water Supply, 7(1), 13–22, doi:10.2166/ws.2007.002, 2007.

Kuhn, T. S.: The Structure of Scientific Revolutions, 1962.

Kumar, P.: Variability, Feedback, and Cooperative Process Dynamics: Elements of a Unifying Hydrologic Theory, Geogr. Compass, 1(6), 1338–1360, doi:10.1111/j.1749-8198.2007.00068.x, 2007.

Kuras, O., Pritchard, J. D., Meldrum, P. I., Chambers, J. E., Wilkinson, P. B., Ogilvy, R. D. and Wealthall, G. P.: Monitoring hydraulic processes with automated time-lapse electrical resistivity tomography (ALERT), Comptes Rendus Geosci., 341(10), 868–885, doi:10.1016/j.crte.2009.07.010, 2009.

Kutbay, U. and Hardalaç, F.: Development of a multiprobe electrical resistivity tomography prototype system and robust underground clustering, Expert Syst., 34(3), e12206, doi:10.1111/exsy.12206, 2017.

Labat, D., Ababou, R. and Mangin, A.: Rainfall–runoff relations for karstic springs. Part I: convolution and spectral analyses, J. Hydrol., 238(3–4), 123–148, doi:10.1016/S0022-1694(00)00321-8, 2000.

Lakoff, G.: Why it Matters How We Frame the Environment, Environ. Commun., 4(1), 70–81, doi:10.1080/17524030903529749, 2010.

Lall, U.: Debates—The future of hydrological sciences: A (common) path forward? One water. One world. Many climes. Many souls, Water Resour. Res., 50(6), 5335–5341, doi:10.1002/2014WR015402, 2014.

Lamb, R. and Beven, K.: Using interactive recession curve analysis to specify a general catchment storage model, Hydrol Earth Syst Sci, 1(1), 101–113, doi:10.5194/hess-1-101-1997, 1997.

Laplace, P.-S.: Essai philosophique sur les probabilités., 1814.

Larocque, M., Mangin, A., Razack, M. and Banton, O.: Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France), J. Hydrol., 205(3–4), 217–231, doi:10.1016/S0022-1694(97)00155-8, 1998.

Li, J., Yuan, D., Liu, J., Jiang, Y., Chen, Y., Hsu, K. L. and Sorooshian, S.: Predicting floods in a large karst river basin by coupling PERSIANN-CCS QPEs with a physically based distributed hydrological model, Hydrol. Earth Syst. Sci., 23(3), 1505–1532, doi:https://doi.org/10.5194/hess-23-1505-2019, 2019.

Liao, T.: Clustering of time series data—a survey, Pattern Recognit., 38(11), 1857–1874, doi:10.1016/j.patcog.2005.01.025, 2005.

Lins, H. F. and Cohn, T. A.: Stationarity: Wanted Dead or Alive?1, JAWRA J. Am. Water Resour. Assoc., 47(3), 475–480, doi:10.1111/j.1752-1688.2011.00542.x, 2011.

Ljung, G. M. and Box, G. E. P.: On a measure of lack of fit in time series models, Biometrika, 65(2), 297–303, doi:10.1093/biomet/65.2.297, 1978.

Loke, M. H. and Barker, R. D.: Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method1, Geophys. Prospect., 44(1), 131–152, doi:10.1111/j.1365-2478.1996.tb00142.x, 1996.

Lorenz, E. N.: Deterministic Nonperiodic Flow, J. Atmospheric Sci., 20(2), 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963.

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U. and Zehe, E.: On the dynamic nature of hydrological similarity, Hydrol. Earth Syst. Sci., 22(7), 3663–3684, doi:https://doi.org/10.5194/hess-22-3663-2018, 2018.

Lutz, S. R., Popp, A., Emmerik, T. van, Gleeson, T., Kalaugher, L., Möbius, K., Mudde, T., Walton, B., Hut, R., Savenije, H., Slater, L. J., Solcerova, A., Stoof, C. R. and Zink, M.: HESS Opinions: Science in today's media landscape – challenges and lessons from hydrologists and journalists, Hydrol. Earth Syst. Sci., 22(7), 3589–3599, doi:https://doi.org/10.5194/hess-22-3589-2018, 2018.

Mackie, J. L.: Causes and Conditions, Am. Philos. Q., 2(4), 245–264, 1965.

Mandelbrot, B. B.: Comment on 'Stochastic Models in Hydrology' by Adrian E. Scheidegger, Water Resour. Res., 6(6), 1791–1791, doi:10.1029/WR006i006p01791, 1970.

Mantovan, P. and Todini, E.: Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, J. Hydrol., 330(1), 368–381, doi:10.1016/j.jhydrol.2006.04.046, 2006.

Marmodoro, A.: Potentiality in Aristotle's Metaphysics, in Handbook of Potentiality, edited by K. Engelhard and M. Quante, pp. 15–43, Springer Netherlands, Dordrecht., 2018.

Matalas, N. C.: Comment on the Announced Death of Stationarity, J. Water Resour. Plan. Manag., 138(4), 311–312, doi:10.1061/(ASCE)WR.1943-5452.0000215, 2012.

Mathevet, T., Lepiller, M. l and Mangin, A.: Application of time-series analyses to the hydrological functioning of an Alpine karstic system: the case of Bange-L'Eau-Morte, Hydrol. Earth Syst. Sci., 8(6), 1051–1064, doi:https://doi.org/10.5194/hess-8-1051-2004, 2004.

Mayaud, C., Wagner, T., Benischke, R. and Birk, S.: Single event time series analysis in a binary karst catchment evaluated using a groundwater model (Lurbach system, Austria), J. Hydrol., 511, 628–639, doi:10.1016/j.jhydrol.2014.02.024, 2014.

McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., et al.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resour. Res., 43(7), doi:10.1029/2006WR005467, 2007.

McGill, R., Tukey, J. W. and Larsen, W. A.: Variations of Box Plots, Am. Stat., 32(1), 12, doi:10.2307/2683468, 1978.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M. and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, Hydrol. Process., 25(4), 511–522, doi:10.1002/hyp.7841, 2011.

Medina, M., Huffaker, R., Jawitz, J. W. and Muñoz-Carpena, R.: Nonlinear Dynamics in Treatment Wetlands: Identifying Systematic Drivers of Nonequilibrium Outlet Concentrations in Everglades STAs, Water Resour. Res., 2018WR024427, doi:10.1029/2018WR024427, 2019.

Melamed, Y. Y. and Lin, M.: Principle of Sufficient Reason, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/spr2018/entries/sufficient-reason/ (Accessed 3 December 2019), 2018.

Menzies, P. and Beebee, H.: Counterfactual Theories of Causation, [online] Available from: https://stanford.library.sydney.edu.au/entries/causation-counterfactual/ (Accessed 24 March 2020), 2001.

Meyfroidt, P.: Approaches and terminology for causal analysis in land systems science, J. Land Use Sci., 11(5), 501–522, doi:10.1080/1747423X.2015.1117530, 2016.

Michel, G., Thys, G. and Commission wallonne d'étude et de protection des sites souterrains: Atlas du Karst Wallon: inventaire cartographique et descriptif des sites karstiques et des circulations d'eau souterraine : Bassine de la Lesse Calestienne., 2015.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P. and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319(5863), 573–574, doi:10.1126/science.1151915, 2008.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., Stouffer, R. J., et al.: On Critiques of "Stationarity is Dead: Whither Water Management?," Rev. Geophys., 7785–7789, doi:10.1002/2015WR017408@10.1002/(ISSN)1944-9208.COMHES1, 2018.

Molini, A., Katul, G. G. and Porporato, A.: Causality across rainfall time scales revealed by continuous wavelet transforms, J. Geophys. Res., 115(D14), D14123, doi:10.1029/2009JD013016, 2010.

Montanari, A.: What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology, Hydrol. Process., 21(6), 841–845, doi:10.1002/hyp.6623, 2007.

Montanari, A. and Koutsoyiannis, D.: A blueprint for process-based modeling of uncertain hydrological systems, Water Resour. Res., 48(9), doi:10.1029/2011WR011412, 2012.

Montanari, A. and Koutsoyiannis, D.: Modeling and mitigating natural hazards: Stationarity is immortal!, Water Resour. Res., 50(12), 9748–9756, doi:10.1002/2014WR016092, 2014.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., et al.: "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, Hydrol. Sci. J., 58(6), 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

Monti, S., Tamayo, P., Mesirov, J. and Golub, T.: Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data, Mach. Learn., 52(1), 91–118, doi:10.1023/A:1023949509487, 2003.

Nash, J. E.: Systematic determination of unit hydrograph parameters, J. Geophys. Res. 1896-1977, 64(1), 111–115, doi:10.1029/JZ064i001p00111, 1959.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W. and Weijs, S. V.: A philosophical basis for hydrological uncertainty, Hydrol. Sci. J., 61(9), 1666–1678, doi:10.1080/02626667.2016.1183009, 2016.

Nguyen, F., Kemna, A., Robert, T. and Hermans, T.: Data-driven selection of the minimum-gradient support parameter in time-lapse focused electric imaging, GEOPHYSICS, 81(1), A1–A5, doi:10.1190/geo2015-0226.1, 2016.

Nimmo, J. R.: Vadose Water, in Encyclopedia of Inland Waters, pp. 766–777, Elsevier., 2009.

Nuzzo, R.: How scientists fool themselves – and how they can stop, Nat. News, 526(7572), 182, doi:10.1038/526182a, 2015.

Ollivier, C., Danquigny, C., Mazzilli, N. and Barbel-Perineau, A.: Contribution of Hydrogeological Time Series Statistical Analysis to the Study of Karst Unsaturated Zone (Rustrel, France), in Hydrogeological and Environmental Investigations in Karst Systems, edited by B. Andreo, F. Carrasco, J. J. Durán, P. Jiménez, and J. W. LaMoreaux, pp. 27–33, Springer Berlin Heidelberg, Berlin, Heidelberg., 2015.

Paasche, H. and Tronicke, J.: Cooperative inversion of 2D geophysical data sets: A zonal approach based on fuzzy c-means cluster analysis, GEOPHYSICS, 72(3), A35–A39, doi:10.1190/1.2670341, 2007.

Paasche, H., Tronicke, J., Holliger, K., Green, A. G. and Maurer, H.: Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy c -means cluster analyses, GEOPHYSICS, 71(3), H33–H44, doi:10.1190/1.2192927, 2006.

Packard, N., Crutchfield, J. P. and S. Shaw, R.: Geometry From a Time Series, Phys Rev Lett, 45(9), 712–716, doi:10.1103/PhysRevLett.45.712, 1980.

Padilla, A. and Pulido-Bosch, A.: Study of hydrographs of karstic aquifers by means of correlation and cross-spectral analysis, J. Hydrol., 168(1–4), 73–89, doi:10.1016/0022-1694(94)02648-U, 1995.

Pagliero, L., Bouraoui, F., Diels, J., Willems, P. and McIntyre, N.: Investigating regionalization techniques for large-scale hydrological modelling, J. Hydrol., 570, 220–235, doi:10.1016/j.jhydrol.2018.12.071, 2019.

Pappenberger, F. and Beven, K. J.: Ignorance is bliss: Or seven reasons not to use uncertainty analysis, Water Resour. Res., 42(5), doi:10.1029/2005WR004820, 2006.

de Pasquale, G., Linde, N., Doetsch, J. and Holbrook, W. S.: Probabilistic inference of subsurface heterogeneity and interface geometry using geophysical data, Geophys. J. Int., 217(2), 816–831, doi:10.1093/gji/ggz055, 2019.

Pearl, J.: Causality: models, reasoning, and inference, Cambridge University Press, Cambridge, U.K. ; New York., 2000.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Peirce, C. S.: Collected Papers of Charles Sanders Peirce, Harvard University Press., 1960.

Perrault, P.: De l'Origine des fontaines, P. le Petit. [online] Available from: https://books.google.be/books?id=EwQzkdcO0i8C, 1674.

Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E. C., Emmerik, T. van, Uijlenhoet, R., Achieng, K., Franz, T. E. and Woods, R.: Scaling, similarity, and the fourth paradigm for hydrology, Hydrol. Earth Syst. Sci., 21(7), 3701–3713, doi:https://doi.org/10.5194/hess-21-3701-2017, 2017.

Pfister, L. and Kirchner, J. W.: Debates—Hypothesis testing in hydrology: Theory and practice, Water Resour. Res., 53(3), 1792–1798, doi:10.1002/2016WR020116, 2017.

Poincaré, H.: Science et méthode, Flammarion., 1908.

Polanyi, M. and Sen, A.: The Tacit Dimension.–, 1966.

Popper, K. R.: The logic of scientific discovery, 1959.

Poulain, A.: Flow and transport characterization in vadose and phreatic zones of karst aquifers: Experimental approaches in the Givetian limestones of South Belgium, University of Namur, Namur. [online] Available from: https://researchportal.unamur.be/en/studentTheses/flow-and-transport-characterization-in-vadose-and-phreatic-zones-, 2017.

Poulain, A., Watlet, A., Kaufmann, O., Camp, M. V., Jourde, H., Mazzilli, N., Rochez, G., Deleu, R., Quinif, Y. and Hallet, V.: Assessment of groundwater recharge processes through karst vadose zone by cave percolation monitoring, Hydrol. Process., 32(13), 2069–2083, doi:10.1002/hyp.13138, 2018.

Quinif, Y., Van Ruymbeke, M., Camelbeeck, T. and Vandycke, S.: Les failles actives de la Grotte de Rochefort (Ardenne, Belgique) sont-elles sismogéniques? Installation d'un laboratoire souterrain, vol. 8, pp. 153–155, Belgium., 1997.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J. and van der Keur, P.: A framework for dealing with uncertainty due to model structure error, Adv. Water Resour., 29(11), 1586–1597, doi:10.1016/j.advwatres.2005.11.013, 2006.

Reggiani, P. and Schellekens, J.: Modelling of hydrological responses: the representative elementary watershed approach as an alternative blueprint for watershed modelling, Hydrol. Process., 17(18), 3785–3789, doi:10.1002/hyp.5167, 2003.

Reichenbach, H.: The Direction of Time, University of California Press. [online] Available from: https://books.google.fr/books?id=f6kNAQAAIAAJ, 1956.

Rinderer, M., Ali, G. and Larsen, L. G.: Assessing structural, functional and effective hydrologic connectivity with brain neuroscience methods: State-of-the-art and research directions, Earth-Sci. Rev., 178, 29–47, doi:10.1016/j.earscirev.2018.01.009, 2018.

Robinson, H.: Dualism, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/fall2017/entries/dualism/ (Accessed 9 December 2019), 2017.

Rodriguez-Iturbe, I., Entekhabi, D. and Bras, R. L.: Nonlinear Dynamics of Soil Moisture at Climate Scales: 1. Stochastic Analysis, Water Resour. Res., 27(8), 1899–1906, doi:10.1029/91WR01035, 1991.

Roobavannan, M., Emmerik, T. H. M. van, Elshafei, Y., Kandasamy, J., Sanderson, M. R., Vigneswaran, S., Pande, S. and Sivapalan, M.: Norms and values in sociohydrological models, Hydrol. Earth Syst. Sci., 22(2), 1337–1349, doi:https://doi.org/10.5194/hess-22-1337-2018, 2018.

Roques, C., Rupp, D. E. and Selker, J. S.: Improved streamflow recession parameter estimation with attention to calculation of − dQ/dt, Adv. Water Resour., 108, 29–43, doi:10.1016/j.advwatres.2017.07.013, 2017.

Rosenblueth, A., Wiener, N. and Bigelow, J.: Behavior, Purpose and Teleology, Philos. Sci., 10(1), 18–24, 1943.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math., 20, 53–65, doi:10.1016/0377-0427(87)90125-7, 1987.

Rücker, C., Günther, T. and Spitzer, K.: Three-dimensional modelling and inversion of dc resistivity data incorporating topography — I. Modelling, Geophys. J. Int., 166(2), 495–505, doi:10.1111/j.1365-246X.2006.03010.x, 2006.

Ruddell, B. L. and Kumar, P.: Ecohydrologic process networks: 1. Identification, Water Resour. Res., 45(3), doi:10.1029/2008WR007279, 2009.

Runge, J.: Causal network reconstruction from time series: From theoretical assumptions to practical estimation, Chaos Interdiscip. J. Nonlinear Sci., 28(7), 075310, doi:10.1063/1.5025050, 2018a.

Runge, J.: Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information, in International Conference on Artificial Intelligence and Statistics, pp. 938–947. [online] Available from: http://proceedings.mlr.press/v84/runge18a.html (Accessed 6 May 2019b), 2018.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D.: Detecting and quantifying causal associations in large nonlinear time series datasets, Sci. Adv., 5(11), eaau4996, doi:10.1126/sciadv.aau4996, 2019a.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., Nes, E. H. van, Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K. and Zscheischler, J.: Inferring causation from time series in Earth system sciences, Nat. Commun., 10(1), 2553, doi:10.1038/s41467-019-10105-3, 2019b.

Rupp, D. E. and Selker, J. S.: Drainage of a horizontal Boussinesq aquifer with a power law hydraulic conductivity profile, Water Resour. Res., 41(11), doi:10.1029/2005WR004241, 2005.

Rupp, D. E. and Selker, J. S.: Information, artifacts, and noise in dQ/dt−Q recession analysis, Adv. Water Resour., 29(2), 154–160, doi:10.1016/j.advwatres.2005.03.019, 2006.

Russell, B.: On the Notion of Cause, Proc. Aristot. Soc., 13, 1–26, 1912.

Saltelli, A., Ed.: Sensitivity analysis in practice: a guide to assessing scientific models, Wiley, Hoboken, NJ., 2004.

Saltelli, A., Chan, K. and Scott, E. M., Eds.: Sensitivity analysis, Wiley, Chichester ; New York., 2000.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S.: Global Sensitivity Analysis. The Primer, John Wiley & Sons, Ltd, Chichester, UK., 2007.

Saltelli, A. and Funtowicz, S.: When all models are wrong: More stringent quality criteria are needed for models used at the science-policy interface, Issues Sci. Technol., 79–85, 2014.

Saltelli, A. and Funtowicz, S.: What is science's crisis really about?, Futures, 91, 5–11, doi:10.1016/j.futures.2017.05.010, 2017.

Salvucci, G. D., Saleem, J. A. and Kaufmann, R.: Investigating soil moisture feedbacks on precipitation with tests of Granger causality, Adv. Water Resour., 25(8), 1305–1312, doi:10.1016/S0309-1708(02)00057-X, 2002.

Samouëlian, A., Cousin, I., Tabbagh, A., Bruand, A. and Richard, G.: Electrical resistivity survey in soil science: a review, Soil Tillage Res., 83(2), 173–193, doi:10.1016/j.still.2004.10.004, 2005.

Sauer, T., Yorke, J. A. and Casdagli, M.: Embedology, J. Stat. Phys., 65(3–4), 579–616, doi:10.1007/BF01053745, 1991.

Savenije, H. H. G.: Equifinality, a blessing in disguise?, Hydrol. Process., 15(14), 2835–2838, doi:10.1002/hyp.494, 2001.

Savenije, H. H. G.: HESS Opinions "The art of hydrology"*, Hydrol. Earth Syst. Sci., 13(2), 157–161, doi:https://doi.org/10.5194/hess-13-157-2009, 2009.

Savenije, H. H. G. and Hrachowitz, M.: HESS Opinions Catchments as meta-organisms – a new blueprint for hydrological modelling, Hydrol. Earth Syst. Sci., 21(2), 1107–1116, doi:10.5194/hess-21-1107-2017, 2017.

Savenije, H. H. G., Hoekstra, A. Y. and van der Zaag, P.: Evolving water science in the Anthropocene, Hydrol. Earth Syst. Sci., 18(1), 319–332, doi:10.5194/hess-18-319-2014, 2014.

Scaini, A., Audebert, M., Hissler, C., Fenicia, F., Gourdol, L., Pfister, L. and Beven, K. J.: Velocity and celerity dynamics at plot scale inferred from artificial tracing experiments and time-lapse ERT, J. Hydrol., 546, 28–43, doi:10.1016/j.jhydrol.2016.12.035, 2017.

Schalge, B., Rihani, J., Baroni, G., Erdal, D., Geppert, G., Haefliger, V., Haese, B., Saavedra, P., Neuweiler, I., Hendricks Franssen, H.-J., Ament, F., Attinger, S., Cirpka, O. A., Kollet, S., Kunstmann, H., Vereecken, H. and Simmer, C.: High-Resolution Virtual Catchment Simulations of the Subsurface-Land Surface-Atmosphere System, Hydrol. Earth Syst. Sci. Discuss., 1–44, doi:https://doi.org/10.5194/hess-2016-557, 2016.

Schreiber, T.: Detecting and Analyzing Nonstationarity in a Time Series Using Nonlinear Cross Predictions, Phys. Rev. Lett., 78(5), 843–846, doi:10.1103/PhysRevLett.78.843, 1997.

Schreiber, T.: Measuring Information Transfer, Phys. Rev. Lett., 85(2), 461–464, doi:10.1103/PhysRevLett.85.461, 2000.

Schrödinger, E.: Nature and the Greeks, University Press. [online] Available from: https://books.google.be/books?id=H7sAAAAAMAAJ, 1954.

Schuler, P., Cantoni, È., Duran, L., Johnston, P. and Gill, L.: Using Wavelet Coherence to Characterize Surface Water Infiltration into a Low-Lying Karst Aquifer, Groundwater, gwat.13012, doi:10.1111/gwat.13012, 2020.

Schwarz, G.: Estimating the Dimension of a Model, Ann. Stat., 6(2), 461–464, doi:10.1214/aos/1176344136, 1978.

SCS: Section 4: hydrology, in National Engineering Handbook, Soil Conservation Service., 1972.

Seibert, J., Uhlenbrook, S. and Wagener, T.: Preface "Hydrology education in a changing world," Hydrol. Earth Syst. Sci., 17(4), 1393–1399, doi:10.5194/hess-17-1393-2013, 2013.

Sendrowski, A. and Passalacqua, P.: Process connectivity in a naturally prograding river delta, Water Resour. Res., 53(3), 1841–1863, doi:10.1002/2016WR019768, 2017.

Shannon, C. E.: A Mathematical Theory of Communication, Bell Syst. Tech. J., 27(3), 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x, 1948.

Shaw, S. B. and Riha, S. J.: Examining individual recession events instead of a data cloud: Using a modified interpretation of dQ/dt–Q streamflow recession in glaciated watersheds to better inform models of low flow, J. Hydrol., 434–435, 46–54, doi:10.1016/j.jhydrol.2012.02.034, 2012.

Shaw, S. B., McHardy, T. M. and Riha, S. J.: Evaluating the influence of watershed moisture storage on variations in base flow recession rates during prolonged rain-free periods in

medium-sized catchments in New York and Illinois, USA, Water Resour. Res., 49(9), 6022–6028, doi:10.1002/wrcr.20507, 2013.

Shen, C.: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resour. Res., 54(11), 8558–8593, doi:10.1029/2018WR022643, 2018.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X. and Tsai, W.-P.: HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community, Hydrol. Earth Syst. Sci., 22(11), 5639–5656, doi:https://doi.org/10.5194/hess-22-5639-2018, 2018.

Sherman, L. K.: Streamflow from rainfall by the unit-graph method, [online] Available from: /paper/Streamflow-from-rainfall-by-the-unit-graph-method-Sherman/e18ca0d7be6af90b1c141e70ad5633888b8eccbe (Accessed 19 March 2020a), 1932a.

Sherman, L. K.: The relation of hydrographs of runoff to size and character of drainage-basins, Trans. Am. Geophys. Union, 13(1), 332, doi:10.1029/TR013i001p00332, 1932b.

Shields, C.: Aristotle, [online] Available from: https://plato.stanford.edu/entries/aristotle/#Sci (Accessed 14 March 2020), 2008.

Shields, C.: Aristotle's Psychology, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2016/entries/aristotle-psychology/ (Accessed 15 March 2020), 2016.

Silberstein, R. P.: Hydrological models are so good, do we still need data?, Environ. Model. Softw., 21(9), 1340–1352, doi:10.1016/j.envsoft.2005.04.019, 2006.

Singh, A., Sharma, S. P., Akca, İ. and Baranwal, V. C.: Fuzzy constrained Lp-norm inversion of direct current resistivity data, GEOPHYSICS, 83(1), E11–E24, doi:10.1190/geo2017-0040.1, 2018.

Singh, V. P.: Entropy theory in hydrologic science and engineering, McGraw-Hill Education, New York., 2015.

Singh, V. P. and Frevert, D. K.: Mathematical Models of Large Watershed Hydrology, Water Resources Publications. [online] Available from: https://books.google.be/books?id=abRyo4OLjgYC, 2002a.

Singh, V. P. and Frevert, D. K.: Mathematical Models of Small Watershed Hydrology and Applications, Water Resources Publications. [online] Available from: https://books.google.be/books?id=Dqm8oc1KsnUC, 2002b.

Singha, K., Day-Lewis, F. D., Johnson, T. and Slater, L. D.: Advances in interpretation of subsurface processes with time-lapse electrical imaging: TIME-LAPSE ELECTRICAL IMAGING, Hydrol. Process., 29(6), 1549–1576, doi:10.1002/hyp.10280, 2015.

Sivakumar, B.: Chaos theory in hydrology: important issues and interpretations, J. Hydrol., 227(1), 1–20, doi:10.1016/S0022-1694(99)00186-9, 2000.

Sivakumar, B.: Rainfall dynamics at different temporal scales: A chaotic perspective, Hydrol Earth Syst Sci, 5(4), 645–652, doi:10.5194/hess-5-645-2001, 2001.

Sivakumar, B.: Dominant processes concept, model simplification and classification framework in catchment hydrology, Stoch. Environ. Res. Risk Assess., 22(6), 737–748, doi:10.1007/s00477-007-0183-5, 2008.

Sivakumar, B.: Hydropsychology: the human side of water research, Hydrol. Sci. J., 56(4), 719–732, doi:10.1080/02626667.2011.580281, 2011a.

Sivakumar, B.: Water crisis: From conflict to cooperation—an overview, Hydrol. Sci. J., 56(4), 531–552, doi:10.1080/02626667.2011.580747, 2011b.

Sivakumar, B.: Chaos in Hydrology, Springer Netherlands, Dordrecht., 2017a.

Sivakumar, B.: Characteristics of Hydrologic Systems, in Chaos in Hydrology: Bridging Determinism and Stochasticity, edited by B. Sivakumar, pp. 29–62, Springer Netherlands, Dordrecht., 2017b.

Sivakumar, B. and Singh, V. P.: Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework, Hydrol. Earth Syst. Sci., 16(11), 4119–4131, doi:10.5194/hess-16-4119-2012, 2012.

Sivapalan, M.: Prediction in ungauged basins: a grand challenge for theoretical hydrology, Hydrol. Process., 17(15), 3163–3170, doi:10.1002/hyp.5155, 2003.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in Encyclopedia of Hydrological Sciences, American Cancer Society., 2006.

Sivapalan, M.: The secret to 'doing better hydrological science': change the question!, Hydrol. Process., 23(9), 1391–1396, doi:10.1002/hyp.7242, 2009.

Sivapalan, M.: From engineering hydrology to Earth system science: milestones in the transformation of hydrologic science, Hydrol Earth Syst Sci, 22(3), 1665–1693, doi:10.5194/hess-22-1665-2018, 2018.

Sivapalan, M. and Blöschl, G.: The Growth of Hydrological Understanding: Technologies, Ideas, and Societal Needs Shape the Field, Water Resour. Res., 53(10), 8137–8146, doi:10.1002/2017WR021396, 2017.

Sivapalan, M., Zhang, L., Vertessy, R. and Blöschl, G.: Downward approach to hydrological prediction, Hydrol. Process., 17(11), 2099–2099, doi:10.1002/hyp.1426, 2003.

Sivapalan, M., Savenije, H. H. G. and Blöschl, G.: Socio-hydrology: A new science of people and water, Hydrol. Process., 26(8), 1270–1276, doi:10.1002/hyp.8426, 2012.

Slovic, P.: Perception of risk, Science, 236(4799), 280–285, doi:10.1126/science.3563507, 1987.

Smakhtin, V. U.: Low flow hydrology: a review, J. Hydrol., 240(3), 147–186, doi:10.1016/S0022-1694(00)00340-1, 2001.

Smith, D. W.: Phenomenology, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/sum2018/entries/phenomenology/ (Accessed 10 December 2019), 2018.

Smuts, J. C.: Holism and evolution, Рипол Классик., 1926.

Sobol', I. M.: On sensitivity estimation for nonlinear mathematical models, Mat. Model., 2(1), 112–118, 1990.

Sood, A. and Smakhtin, V.: Global hydrological models: a review, Hydrol. Sci. J., 60(4), 549–565, doi:10.1080/02626667.2014.950580, 2015.

Spade, P. V.: Medieval Philosophy, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/sum2018/entries/medieval-philosophy/ (Accessed 18 February 2020), 2018.

Spade, P. V. and Panaccio, C.: William of Ockham, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online]

Available from: https://plato.stanford.edu/archives/spr2019/entries/ockham/ (Accessed 3 December 2019), 2019.

Spirtes, P. and Glymour, C.: An Algorithm for Fast Recovery of Sparse Causal Graphs, Soc. Sci. Comput. Rev., 9(1), 62–72, doi:10.1177/089443939100900106, 1991.

Spirtes, P., Glymour, C. and Scheines, R.: Causation, Prediction, and Search, Springer-Verlag, New York., 1993.

Srinivasan, V., Lambin, E. F., Gorelick, S. M., Thompson, B. H. and Rozelle, S.: The nature and causes of the global water crisis: Syndromes from a meta-analysis of coupled human-water studies, Water Resour. Res., 48(10), doi:10.1029/2011WR011087, 2012.

Srinivasan, V., Sanderson, M., Garcia, M., Konar, M., Blöschl, G. and Sivapalan, M.: Prediction in a socio-hydrological world, Hydrol. Sci. J., 62(3), 338–345, doi:10.1080/02626667.2016.1253844, 2017.

Stedinger, J. R., Vogel, R. M., Lee, S. U. and Batchelder, R.: Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, Water Resour. Res., 44(12), doi:10.1029/2008WR006822, 2008.

Stern, P. C.: Toward a coherent theory of environmentally significant behavior, J. Soc. Issues, 56(3), 407–424, doi:10.1111/0022-4537.00175, 2000.

Stern, P. C., Dietz, T., Abel, T., Guagnano, G. A. and Kalof, L.: A value-belief-norm theory of support for social movements: The case of environmentalism, Hum. Ecol. Rev., 81–97, 1999.

Stewart, M.K.: Promising new baseflow separation and recession analysis methods applied to streamflow at Glendhu Catchment, New Zealand. Hydrol. Earth Syst. Sci., 19, 2587–2603, doi:10.5194/hess-19-2587-2015, 2015.

Stigler, S. M.: Francis Galton's Account of the Invention of Correlation, Stat. Sci., 4(2), 73–79, doi:10.1214/ss/1177012580, 1989.

Stoelzle, M., Stahl, K. and Weiler, M.: Are streamflow recession characteristics really characteristic?, Hydrol Earth Syst Sci, 17(2), 817–828, doi:10.5194/hess-17-817-2013, 2013.

Sugihara, G.: Nonlinear forecasting for the classification of natural time series, Phil Trans R Soc Lond A, 348(1688), 477–495, doi:10.1098/rsta.1994.0106, 1994.

Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature, 344(6268), 734–741, doi:10.1038/344734a0, 1990.

Sugihara, G., May, R., Ye, H., Hsieh, C. -h., Deyle, E., Fogarty, M. and Munch, S.: Detecting Causality in Complex Ecosystems, Science, 338(6106), 496–500, doi:10.1126/science.1227079, 2012.

Tagne, G. V. and Dowling, C.: Inferring groundwater flow and recharge from time series analysis of storm responses in a karst aquifer of southeastern Kentucky (USA), Hydrogeol. J., 26(8), 2649–2668, doi:10.1007/s10040-018-1837-8, 2018.

Takens, F.: Detecting strange attractors in turbulence, in Dynamical Systems and Turbulence, Warwick 1980, edited by D. Rand and L.-S. Young, pp. 366–381, Springer Berlin Heidelberg., 1981.

Tallaksen, L. M.: A review of baseflow recession analysis, J. Hydrol., 165(1), 349–370, doi:10.1016/0022-1694(94)02540-R, 1995.

Tan, P.-N., Steinbach, M., Kumar, V. and Karpatne, A.: Introduction to Data Mining, Global Edition, Pearson Education Limited, Harlow, United Kingdom. [online] Available from:

https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5720020 (Accessed 19 September 2019), 2019.

Taylor, C. C. W. and Lee, M.-K.: The Sophists, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2016/entries/sophists/ (Accessed 27 January 2020), 2016.

Terpstra, T.: Emotions, Trust, and Perceived Risk: Affective and Cognitive Routes to Flood Preparedness Behavior, Risk Anal., 31(10), 1658–1675, doi:10.1111/j.1539-6924.2011.01616.x, 2011.

Theiler, J.: Spurious dimension from correlation algorithms applied to limited time-series data, Phys. Rev. Gen. Phys., 34(3), 2427–2432, 1986.

Thompson, S. E., Sivapalan, M., Harman, C. J., Srinivasan, V., Hipsey, M. R., Reed, P., Montanari, A. and Blöschl, G.: Developing predictive insight into changing water systems: use-inspired hydrologic science for the Anthropocene, Hydrol. Earth Syst. Sci., 17(12), 5013–5039, doi:https://doi.org/10.5194/hess-17-5013-2013, 2013.

Todini, E.: Hydrological catchment modelling: past, present and future, Hydrol. Earth Syst. Sci., 11(1), 468–482, doi:https://doi.org/10.5194/hess-11-468-2007, 2007.

Triantafyllou, A., Watlet, A., Le Mouélic, S., Camelbeeck, T., Civet, F., Kaufmann, O., Quinif, Y. and Vandycke, S.: 3-D digital outcrop model for analysis of brittle deformation and lithological mapping (Lorette cave, Belgium), J. Struct. Geol., 120, 55–66, doi:10.1016/j.jsg.2019.01.001, 2019.

Troch, P. A., Troch, F. P. D. and Brutsaert, W.: Effective water table depth to describe initial conditions prior to storm rainfall in humid regions, Water Resour. Res., 29(2), 427–434, doi:10.1029/92WR02087, 1993.

Troch, P. A., Berne, A., Bogaart, P., Harman, C., Hilberts, A. G. J., Lyon, S. W., Paniconi, C., et al.: The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange, Water Resour. Res., 49(9), 5099–5116, doi:10.1002/wrcr.20407, 2013.

Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T. and Valdés-Pineda, R.: Catchment coevolution: A useful framework for improving predictions of hydrological change?, Water Resour. Res., 51(7), 4903–4922, doi:10.1002/2015WR017032, 2015.

Tuttle, S. E. and Salvucci, G. D.: Confounding factors in determining causal soil moisture-precipitation feedback, Water Resour. Res., 53(7), 5531–5544, doi:10.1002/2016WR019869, 2017.

Tversky, A. and Kahneman, D.: Judgment under Uncertainty: Heuristics and Biases, Science, 185(4157), 1124–1131, doi:10.1126/science.185.4157.1124, 1974.

Tversky, A. and Kahneman, D.: The framing of decisions and the psychology of choice, Science, 211(4481), 453–458, doi:10.1126/science.7455683, 1981.

Van Camp, M., de Viron, O., Pajot-Métivier, G., Casenave, F., Watlet, A., Dassargues, A. and Vanclooster, M.: Direct measurement of evapotranspiration from a forest using a superconducting gravimeter: EVAPOTRANSPIRATION AND GRAVITY, Geophys. Res. Lett., 43(19), 10,225-10,231, doi:10.1002/2016GL070534, 2016.

Vanclooster, M., Boesten, J. J. T. I., Trevisan, M., Brown, C. D., Capri, E., Eklo, O. M., Gottesbüren, et al.: A European test of pesticide-leaching models: methodology and major recommendations, Agric. Water Manag., 44(1), 1–19, doi:10.1016/S0378-3774(99)00081-5, 2000.

Vejmelka, M. and Paluš, M.: Inferring the directionality of coupling with conditional mutual information, Phys. Rev. E, 77(2), 026214, doi:10.1103/PhysRevE.77.026214, 2008.

Vemuri, V. and Vemuri, N.: On the systems approach in hydrology, Hydrol. Sci. J., 15(2), 17–38, 1970.

Vinh, N. X. and Epps, J.: A Novel Approach for Automatic Number of Clusters Detection in Microarray Data Based on Consensus Clustering, in 2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering, pp. 84–91, IEEE, Taichung, Taiwan., 2009.

Vinh, N. X., Epps, J. and Bailey, J.: Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance, J. Mach. Learn. Res., (11), 2837–2854, 2010.

de Viron, O., Delforge, D. and Ghil, M.: Causal relations in the climate system, in AGU Fall Meeting 2019, AGU., 2019.

Vogel, R. M. and Kroll, C. N.: Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, Water Resour. Res., 28(9), 2451–2458, doi:10.1029/92WR01007, 1992.

Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P. K., Hooper, R. P. and Matalas, N. C.: Hydrology: The interdisciplinary science of water, Water Resour. Res., 51(6), 4409–4430, doi:10.1002/2015WR017049, 2015.

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V. and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Environ. Res. Risk Assess., 23(7), 1011–1026, doi:10.1007/s00477-008-0274-y, 2009.

Wagener, T., Sivapalan, M., Troch, P. and Woods, R.: Catchment Classification and Hydrologic Similarity, Geogr. Compass, 1(4), 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., et al.: The future of hydrology: An evolving science for a changing world, Water Resour. Res., 46(5), doi:10.1029/2009WR008906, 2010.

Wang, Y., Yang, J., Chen, Y., De Maeyer, P., Li, Z. and Duan, W.: Detecting the Causal Effect of Soil Moisture on Precipitation Using Convergent Cross Mapping, Sci. Rep., 8(1), 12171, doi:10.1038/s41598-018-30669-2, 2018.

Ward, J. H.: Hierarchical Grouping to Optimize an Objective Function, J. Am. Stat. Assoc., 58(301), 236–244, doi:10.1080/01621459.1963.10500845, 1963.

Ward, W. O. C., Wilkinson, P. B., Chambers, J. E., Oxby, L. S. and Bai, L.: Distribution-based fuzzy clustering of electrical resistivity tomography images for interface detection, Geophys. J. Int., 197(1), 310–321, doi:10.1093/gji/ggu006, 2014.

Ward, W. O. C., Wilkinson, P. B., Chambers, J. E., Nilsson, H., Kuras, O. and Bai, L.: Tracking tracer motion in a 4-D electrical resistivity tomography experiment: TRACKING TRACER MOTION IN 4-D ERT, Water Resour. Res., 52(5), 4078–4094, doi:10.1002/2015WR017958, 2016.

Watlet, A.: Hydrogeophysical monitoring of groundwater recharge processes through the karst vadose zone at Rochefort (Belgium), University of Mons, Mons, Belgium., 2017.

Watlet, A., Kaufmann, O., Triantafyllou, A., Poulain, A., Chambers, J. E., Meldrum, P. I., Wilkinson, P. B., et al.: Data And Results For Manuscript “Imaging Groundwater Infiltration Dynamics In Karst Vadose Zone With Long-Term Ert Monitoring,” , doi:10.5281/zenodo.1158631, 2018a.

Watlet, A., Kaufmann, O., Triantafyllou, A., Poulain, A., Chambers, J. E., Meldrum, P. I., Wilkinson, P. B., et al.: Imaging groundwater infiltration dynamics in the karst vadose zone with long-term ERT monitoring, Hydrol. Earth Syst. Sci., 22(2), 1563–1592, doi:https://doi.org/10.5194/hess-22-1563-2018, 2018b.

Watlet, A., Camp, M. V., Francis, O., Poulain, A., Rochez, G., Hallet, V., Quinif, Y. and Kaufmann, O.: Gravity Monitoring of Underground Flash Flood Events to Study Their Impact on Groundwater Recharge and the Distribution of Karst Voids, Water Resour. Res., 56(4), e2019WR026673, doi:10.1029/2019WR026673, 2020.

Weijs, S. V. and Ruddell, B. L.: Debates: Does Information Theory Provide a New Paradigm for Earth Science? Sharper Predictions Using Occam's Digital Razor, Water Resour. Res., 56(2), e2019WR026471, doi:10.1029/2019WR026471, 2020.

Weiler, M. and Beven, K.: Do we need a Community Hydrological Model?, Water Resour. Res., 51(9), 7777–7784, doi:10.1002/2014WR016731, 2015.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S. and Rosenbaum, R. M.: Perceiving the causes of success and failure, in Attribution: Perceiving the causes of behavior, pp. 95–120, Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US., 1987.

Wheeler, S. C.: Plato's Enlightenment: The Good as the Sun, Hist. Philos. Q., 14(2), 171–188, 1997.

White, G.: Medieval Theories of Causation, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/sum2018/entries/causation-medieval/ (Accessed 18 February 2020), 2018.

White, W.: Groundwater Flow in Karstic Aquifers, in The Handbook of Groundwater Engineering, Second Edition, pp. 21-1-21–47, CRC Press. [online] Available from: http://dx.doi.org/10.1201/9781420006001.ch21 (Accessed 17 January 2017), 2006.

Wiener, N.: Cybernetics; or control and communication in the animal and the machine., 1948.

Wiener, N.: The theory of prediction, Mod. Math. Eng., 1956.

Wilkinson, P., Chambers, J., Meldrum, P., Watson, C., Inauen, C., Swift, R. and Curioni, G.: The Automated Geoelectrical Data Processing Workflow of the PRIME Infrastructure Monitoring System., 2019.

Wilson, G. V., Nieber, J. L., Fox, G. A., Dabney, S. M., Ursic, M. and Rigby, J. R.: Hydrologic connectivity and threshold behavior of hillslopes with fragipans and soil pipe networks, Hydrol. Process., 31(13), 2477–2496, doi:10.1002/hyp.11212, 2017.

Wittenberg, H.: Nonlinear analysis of flow recession curves, IAHS Publ.-Ser. Proc. Rep.-Intern Assoc Hydrol. Sci., 221, 61–68, 1994.

Wittenberg, H.: Baseflow recession and recharge as nonlinear storage processes, Hydrol. Process., 13(5), 715–726, 1999.

Wittenberg, H. and Sivapalan, M.: Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation, J. Hydrol., 219(1), 20–33, doi:10.1016/S0022-1694(99)00040-2, 1999.

WMO: International Glossary of Hydrology (WMO- No. 385)., 2012.

Wolfsdorf, D.: Plato's Conception of Knowledge, Class. World, 105(1), 57–75, 2011.

Wood, E. F., Roundy, J. K., Troy, T. J., Beek, L. P. H. van, Bierkens, M. F. P., Blyth, E., Roo, et al.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47(5), doi:10.1029/2010WR010090, 2011.

Woods, R.: Seeing catchments with new eyes. Spatial patterns in catchment hydrology: observations and modelling Rodger Grayson, Günter Blöschl (Eds.) Cambridge University Press, 416 pp. GBP 65.00, USD$95, AUD215 (ISBN 0-521-63316-8) Published 2000, Hydrol. Process., 16(5), 1111–1113, doi:10.1002/hyp.539, 2002.

Woodward, J.: Causation and Manipulability, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. [online] Available from: https://plato.stanford.edu/archives/win2016/entries/causation-mani/ (Accessed 5 August 2019), 2016.

Wouters, B., Bonin, J. A., Chambers, D. P., Riva, R. E. M., Sasgen, I. and Wahr, J.: GRACE, time-varying gravity, Earth system dynamics and climate change, Rep. Prog. Phys., 77(11), 116801, doi:10.1088/0034-4885/77/11/116801, 2014.

Wright, S.: Correlation and causation, J. Agric. Res., 20(7), 557–585, 1921.

Xu, S., Sirieix, C., Riss, J. and Malaurent, P.: A clustering approach applied to time-lapse ERT interpretation — Case study of Lascaux cave, J. Appl. Geophys., 144, 115–124, doi:10.1016/j.jappgeo.2017.07.006, 2017.

Ye, H. and Sugihara, G.: Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality, Science, 353(6302), 922–925, doi:10.1126/science.aag0863, 2016.

Ye, H., Deyle, E. R., Gilarranz, L. J. and Sugihara, G.: Distinguishing time-delayed causal interactions using convergent cross mapping, Sci. Rep., 5, 14750, doi:10.1038/srep14750, 2015.

Yevjevich, V.: Stochastic models in hydrology, Stoch. Hydrol. Hydraul., 1(1), 17–36, doi:10.1007/BF01543907, 1987.

Young, P. C.: The data-based mechanistic approach to the modelling, forecasting and control of environmental systems, Annu. Rev. Control, 30(2), 169–182, doi:10.1016/j.arcontrol.2006.05.002, 2006.

Young, P. C.: Hypothetico-inductive data-based mechanistic modeling of hydrological systems, Water Resour. Res., 49(2), 915–935, doi:10.1002/wrcr.20068, 2013.

Yule, G. U.: On the theory of correlation for any number of variables, treated by a new system of notation, Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character, 79(529), 182–193, doi:10.1098/rspa.1907.0028, 1907.

Yule, G. U.: On the Time-Correlation Problem, with Especial Reference to the Variate-Difference Correlation Method, J. R. Stat. Soc., 84(4), 497–537, doi:10.2307/2341101, 1921.

Zehe, E. and Blöschl, G.: Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, Water Resour. Res., 40(10), doi:10.1029/2003WR002869, 2004.

Zehe, E. and Sivapalan, M.: Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications, Hydrol. Earth Syst. Sci., 13(7), 1273–1297, doi:10.5194/hess-13-1273-2009, 2009.

Zhou, J., Revil, A., Karaoulis, M., Hale, D., Doetsch, J. and Cuttler, S.: Image-guided inversion of electrical resistivity data, Geophys. J. Int., 197(1), 292–309, doi:10.1093/gji/ggu001, 2014.

Ziman, J. M.: Reliable Knowledge: An Exploration of the Grounds for Belief in Science, Cambridge University Press., 1978.

Appendices

Appendix I. Causality in science: a bibliometric analysis

Causality was introduced in science by Aristotle as a proper way to provide an explanation. In the 18th century, Leibniz argued that anything in the real world could not occur with a sufficient reason or cause. Not much later, the skeptical conclusion of the philosopher David Hume showed that causality could not be demonstrated [1]. What we see, as we see it, is, at best, only a constant association between two observations that we call cause and effect. In 1912, the philosopher and mathematician Bertrand Russel argued that reference to the law of causality had almost disappeared from science; however, surviving among philosophers like a relic of the past [2]. Is this gradual disappearance of causality still undergoing? Today, three centuries after Hume and one after Russell, what is the status of causality in science?

To address this question, this note conducts a bibliometric analysis of the SCOPUS scientific literature database. We report statistics for 33 scientific areas about the uses in title, abstract, and keywords within scientific articles of the words: causality, causative, causal, and causation. The statistics cover the general interest in causality per domain and the current trends computed for the period 1999-2019.

Methodology

To measure the trend in the use of words related to causality in scientific articles, it is essential to account relatively for the trend in the publication of scientific articles. Indeed, a positive trend in the references to causality could be simply the reflection of the increase of scientific publications. Hence, for a given scientific area (**AREA**), two research equations should be performed. The first one focuses on causality related words:

> TITLE-ABS-KEY(causal OR causality OR causative OR causation) AND SUBJAREA(***AREA***) AND DOCTYPE(ar) AND (PUBYEAR > 1998 AND PUBYEAR < 2020)

The term "cause" was avoided because of its broad meaning, and we preferred to refer to the specific four terms. The command DOCTYPE(ar) allows us to retrieve scientific articles only, and PUBYEAR allows us to constraint the

publication period between 1999 and 2019. The second request focuses on the retrieval of all publication in the scientific area (AREA) for that same period:

> SUBJAREA(***AREA***) AND DOCTYPE(ar) AND (PUBYEAR > 1998 AND PUBYEAR < 2020)

Each request is arranged to get the count of publication per year. A general interest can be computed by dividing the number of publications related to causality by the total number of publications for every year. Least square fitting of a linear model allows computing a trend in this general interest and its significance for the 20 years (p-value).

Table I.1 reports all the subject areas that are investigated. The four-letter codes are subject areas that are defined within SCOPUS. The codes with an asterisk were defined by specific research equations. ALL denotes the request when no subject areas are mentioned. The two-letter codes HS, LS, PS, and SS are general requests over the scientific domain obtained with the OR operator (e.g., for Health sciences HS, SUBJAREA(MEDI OR NURS OR VETE OR DENT OR HEAL OR MULT)). HYDRO was not included in the definition of physical science PS. To identify studies related to hydrological sciences HYDRO, the research equation was modified to include the command SRCTITLE(hydro* OR water) such that only articles from a source name including the prefix "hydro" or the word "water" are retrieved. The subject areas in SCOPUS are not mutually exclusive.

Table I.1: Scientific subject area, domain and related code

DOMAIN	SUBJECT AREA	CODE
Life Sciences	Agricultural and Biological Sciences	AGRI
All Sciences	All sciences	ALL*
Social Sciences	Arts and Humanities	ARTS
Life Sciences	Biochemistry, Genetics and Molecular Biology	BIOC
Social Sciences	Business, Management and Accounting	BUSI
Physical Sciences	Chemical Engineering	CENG
Physical Sciences	Chemistry	CHEM
Physical Sciences	Computer Sciences	COMP
Social Sciences	Decision Sciences	DECI
Health Sciences	Dentistry	DENT
Physical Sciences	Earth and Planetary Sciences	EART
Social Sciences	Economics, Econometrics, and Finance	ECON
Physical Sciences	Energy	ENER
Physical Sciences	Engineering	ENGI
Physical Sciences	Environmental Sciences	ENVI
Health Sciences	Health Professions	HEAL
Health Sciences	All health sciences	HS*
Physical Sciences	Hydrological sciences	HYDRO*
Life Sciences	Immunology and Microbiology	IMMU
Life Sciences	All life sciences	LS*
Physical Sciences	Materials Science	MATE
Physical Sciences	Mathematics	MATH
Health Sciences	Medicine	MEDI
Health Sciences	Multidisciplinary	MULT
Life Sciences	Neurosciences	NEUR
Health Sciences	Nursing	NURS
Life Sciences	Pharmacology, Toxicology, and Pharmaceutics	PHAR
Physical Sciences	Physics and Astronomy	PHYS
Physical Sciences	All physical sciences	PS*
Social Sciences	Psychology	PSYC
Social Sciences	Social Sciences	SOCI
Social Sciences	ALL social sciences	SS*
Health Sciences	Veterinary	VETE

*Defined by user

Results

Table I.2 shows the results of the bibliometric analysis per subject area (Table I.1). The results are sorted by the average percentage of causality related papers per year, which denotes a general interest or affinity in the use of causality related terms in the title, abstract, and keywords of articles. Besides, Figure I.1 reports the trends for all publication (ALL), and each scientific

domain: social sciences (SS), health sciences (HS), life sciences (LS), and physical sciences (PS).

Table I.2: Result of the bibliometric analysis per subject area (see Table I.1)

Subject Area	Averaged nbr of publication per year	Averaged nbr of causality related publications per year	Averaged percentage of causality related publications per year	Trend in causality related paper (% per year)	Trend p-value
ECON	27629	557	1.76	0.094	5.00E-12
PSYC	38651	610	1.53	0.021	1.84E-08
IMMU	50692	684	1.28	0.054	9.97E-16
VETE	16043	203	1.20	0.046	3.74E-10
NEUR	43495	545	1.15	0.067	3.41E-16
SS	217637	2479	1.05	0.038	4.57E-16
BUSI	40165	453	0.98	0.069	3.16E-14
MEDI	407902	4099	0.97	0.027	3.66E-17
HS	475558	4709	0.95	0.027	6.97E-16
ARTS	54031	509	0.90	0.020	1.23E-05
SOCI	120708	1150	0.88	0.027	4.88E-14
DECI	13961	133	0.86	0.042	1.89E-10
MULT	29297	311	0.85	0.058	8.40E-10
LS	400600	3389	0.79	0.034	5.13E-17
AGRI	139693	1142	0.76	0.026	3.71E-12
BIOC	208402	1658	0.74	0.037	1.30E-14
NURS	27582	197	0.71	0.006	2.45E-01*
ALL	1504271	9556	0.60	0.020	1.85E-18
HEAL	18369	108	0.57	0.011	7.78E-05
ENVI	88603	494	0.52	0.015	6.88E-10
PHAR	55327	288	0.49	0.019	8.76E-13
DENT	10026	49	0.48	0.003	2.17E-01*
COMP	86817	388	0.43	0.008	3.45E-05
MATH	89035	353	0.37	0.014	3.44E-11
EART	75232	245	0.31	0.009	1.05E-08
PS	729356	1856	0.24	0.009	1.23E-14
HYDRO	10398	22	0.20	0.006	6.62E-04
ENER	44622	113	0.19	0.019	2.35E-11
ENGI	238856	444	0.17	0.006	3.29E-11
PHYS	195776	308	0.15	0.004	5.97E-06
CENG	80092	71	0.08	0.005	5.36E-11
CHEM	169589	123	0.07	0.004	2.02E-11
MATE	167026	70	0.04	0.001	4.71E-05

*Not significant

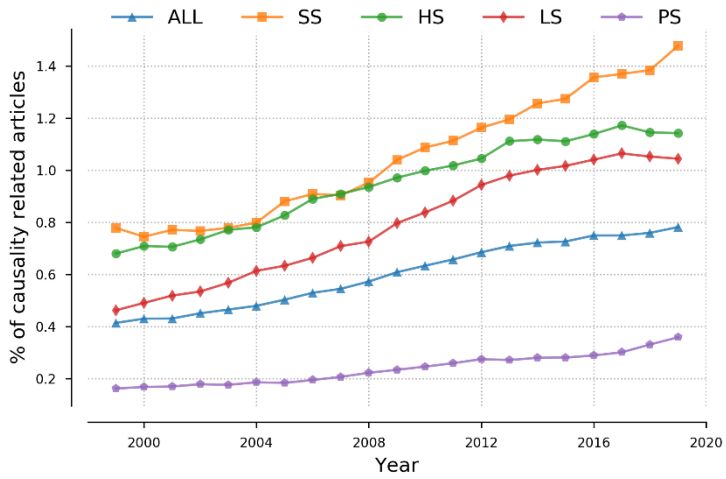

Figure I.1: Trends in causality related terms in the titles, keywords, and abstracts of scientific articles (1999-2019) for all SCOPUS articles (ALL), social sciences (SS), health sciences (HS), life sciences (LS), and physical sciences (PS).

Discussion

What is noteworthy in Table I.2 is the two general trends. First, the use of causal vocabulary increased in all disciplines during the past 20 years. Secondly, the use of causality terms in a discipline seems to be positively correlated with the complexity of the systems studied by that discipline, in that order: social sciences health sciences, life sciences, physical sciences (see also Figure I.1).

Hence, Social sciences dominate the ranking, strongly supported by publications in Economics, Econometrics, and Finance (ECON), that exhibit the strongest interest and the highest trend. This illustrates the strong motivation to understand markets and the rules that govern them by the invisible hand of Adam Smith. This strong interest can be traced back to the pioneering work of economist and Nobel laureate Granger and its work to assess causality from time series using vector autoregressive models [3]. From that perspective, causality is defined by the predictive power of a past and prior variable, the cause, over a present one, the effect. The interest in

causality in social sciences is also supported by studies in psychology. This causal perspective is quite different from considering that causality is perceived differently by individuals in phase with Hume's conclusions [1]. Psychological studies are therefore usually based or rooted in the causal attribution theory, i.e., on how individuals attribute causes of events to people's behaviors or to their external environment. The causal attribution theory goes back to Michotte, Heider, or Kelley [4–6], and has percolated more broadly in social sciences (SOCI) as causal attribution may depend on socio-cultural factors. It further studies cognitive biases, which are also of interest to address scientific biases.

Health Sciences (SS) shows the second rank in their interest in causality, primarily dominated by publication related to the vast subject area of medicine (MEDI). The interest of medicine in causality has been proven since Hippocrates in Ancient Greece, who is credited with laying the foundation of modern etiology in medicine, the study of the causes of diseases. Since then, in medicine, all diseases shall have a natural cause and not a religious or mystical one. This influence is also perceived in immunology and microbiology (IMMU), which are classified within life sciences (LS). Neurosciences is a particular domain were time-series analysis techniques, such as Granger causality, are often applied to understand neuronal stimuli, but it is also deeply related to the question of free will and consciousness, that are central philosophical concepts in the causality debate [7].

Life sciences occupy third place overall. After IMMU, Agricultural and Biological Sciences (AGRI), followed by biochemistry, genetics, and molecular Biology (BIOC), are above average in the use of the term associated with causality. Causality was indeed reintroduced by Wright that had developed one of the first applications of causal graphs on agrometeorological variables [8]. The biologist Bertalanffy introduced his general system theory based on an Aristotelean definition of causality [9]. It enriched causality from the simple notion of the prior cause, as a complex system should also have a form of organization that fits a purpose, i.e., a final cause. Since then, two kinds of causality are acknowledged based on the notion of control. Bottom-up causality refers to an explanation from the control of the parts on the whole, while top-down causality explains the control of the whole system on its parts. The general system theory has also influenced social sciences or any discipline that relates to systems that have components with different functions and

exhibit organization, emergent properties, stability, or resilience in its form or function.

Still, Pharmacology, Toxicology, and Pharmaceutics (PHAR) are below the averaged. The same holds for most of the physical science disciplines (PS) at the bottom of Table I.2, eventually because they are more focused on a linear scheme of causation: the action-reaction paradigm. It may also reflect this attitude of retreat behind facts, the scientific method, the logic of reasoning, the very one that pushed Russell to banish causality from science, and that is best pictured by the words of Newton: “I frame no hypothesis”. It must be noticed that this reluctant attitude in providing causal explanations may be inherited from history since Descartes advocated for rationalism in response to the dogmatic prosecution of Galileo by the religious instances, further encouraged by Hume and its conclusion that causality is not purely rational. A final potential reason on the disinclination to use causality related terms in PS, is perhaps the tautology that exists between the word “physical” and “causal”. It is indeed redundant to speak of a physical causal relationship. In most case, what is implicitly meant with a physical relationship is a causal relationship, and causality is an unnecessary duplicate of physics.

Within PS, environmental sciences (ENVI) are the top referrer to the causal terminology as it treats with more complex systems, relies on more uncertain physics, and may have been influenced more profoundly by AGRI. Despite a relative reluctance in speaking openly about causality, Mathematics, Computer Sciences, Engineering, or Physics, provided significant advances in the study of causality. This is notably the case of Wiener's work, which proposed a systemic vision of science, cybernetics [10], inspired by guidance, steering, and navigation systems in engineering (i.e., guided with purpose), based on information theory developed in parallel by Shannon [11] and whose mathematics and terminology are close to those used in statistical mechanics and thermodynamics (e.g., information entropy). His theory was integrated into Bertalanffy’s general system theory. Besides, Wiener's definition of causality based on predictability [12] was the one considered by Granger in his linear vector autoregressive framework for causality [3]. In the meantime, plenty of methods have been developed for causal inference that may be applied many fields, based on the pioneering work of Wright on the probabilistic causal graph [13]–[15], or on the Weiner-Granger time-series analysis framework by accommodating it to nonlinear dynamics [16], [17]. The theory of nonlinear dynamical systems or chaos theory introduced by the

mathematician Poincarré also laid to the development of recent causal inference techniques from time-series [18].

Conclusion

This note analyzed the occurrence of causality related terms (causality, causal, causative, and causation) in the most vital parts of scientific articles to convey messages: titles, keywords, and abstracts. The bibliometric analysis is based on the SCOPUS database and on 33 scientific domains or subject areas (sub-domains). The terms are observed on average in 0.6 % of scientific publications. All scientific areas exhibit a positive trend in the usage of causality related terms along the past 20 years, on average of + 0.02% per year, indicating a potential come-back of causality in the vocabulary of sciences. In general, scientific domains that study complex systems (e.g., Social Sciences, Health Sciences, Life sciences) use more frequently such terminology (respectively, on average 1.05%, 0.95%, 0.79% with a trend of +0.038%, +0.027%; +0.034% per year). The usage is relatively timorous in Physical sciences (0.24% with a trend of +0.009% per year), especially disciplines studying simple mechanics based on the action-reaction paradigm, despite the critical developments that were made in mathematics, computer sciences, or engineering to address causal inference.

References

[1] D. Hume, Philosophical Essays Concerning Human Understanding. A. Millar, 1748.
[2] B. Russell, "On the Notion of Cause," Proc. Aristot. Soc., vol. 13, pp. 1–26, 1912.
[3] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," Econometrica, vol. 37, no. 3, pp. 424–438, 1969, doi: 10.2307/1912791.
[4] F. Heider, The psychology of interpersonal relations. Hoboken, NJ, US: John Wiley & Sons Inc, 1958.
[5] H. H. Kelley, "The processes of causal attribution," Am. Psychol., vol. 28, no. 2, pp. 107–128, 1973, doi: 10.1037/h0034225.
[6] A. Michotte, La perception de la causalité. Editions de l'Institut Supérieur de Philosophie, 1946.
[7] B. Feltz, M. Missal, and A. Sims, Eds., Free will, causality, and neuroscience. Leiden: Brill Rodopi, 2020.
[8] S. Wright, "Correlation and causation," J. Agric. Res., vol. 20, no. 7, pp. 557–585, 1921.
[9] L. von Bertalanffy, General system theory: foundations, development, applications. G. Braziller, 1968.
[10] N. Wiener, "Cybernetics; or control and communication in the animal and the machine.," 1948.

[11] C. E. Shannon, "A Mathematical Theory of Communication," Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
[12] N. Wiener, "The theory of prediction," Mod. Math. Eng., 1956.
[13] J. Pearl, Causality: models, reasoning, and inference. Cambridge, U.K. ; New York: Cambridge University Press, 2000.
[14] P. Spirtes, C. Glymour, and R. Scheines, Causation, Prediction, and Search. New York: Springer-Verlag, 1993.
[15] C. N. Glymour, The mind's arrows: Bayes nets and graphical causal models in psychology. Cambridge, Mass: MIT Press, 2001.
[16] T. Schreiber, "Measuring Information Transfer," Phys. Rev. Lett., vol. 85, no. 2, pp. 461–464, Jul. 2000, doi: 10.1103/PhysRevLett.85.461.
[17] J. Runge et al., "Inferring causation from time series in Earth system sciences," Nat. Commun., vol. 10, no. 1, p. 2553, Jun. 2019, doi: 10.1038/s41467-019-10105-3.
[18] G. Sugihara et al., "Detecting Causality in Complex Ecosystems," Science, vol. 338, no. 6106, pp. 496–500, Oct. 2012, doi: 10.1126/science.1227079.

Appendix II. MIGRADAKH database description

The MIGRADAKH database consists of time series data, typically time and value paired observations sampled at a fixed frequency, delivered in text format following a set of fixed rules regarding: files naming, files structure, number format, missing values, as well as conventions about date, time (UTC+0) and frequencies. The encoding is programmatically assisted and provides control to assert that file structure is correct, together with a systematic check of values to report either abnormal values or duplicated entries. In addition, each time-series data is associated with a description in a SQL metadatabase. The metadatabase follows a simplified structure of the WaterML 2 data model. It allows describing, for each time series, the observed property, the type of sensors, the sampled medium, on-site interventions, the people or institution involved at each step of the data generation procedure, and more detailed descriptions of the data generation process and its quality. Therefore, the metadata ware gathered from available information are provisional, unequal, and need full validation.

MIGRADAKH database consists of a CSV file database, 1 per time-series, a SQL metadatabase, and a pdf catalog allowing to have a quick look at the time-series, the location of its sensor, and summary statistics. The CSV files and the SQL metadatabase are archived at UCLouvain. The file database respect the following file naming convention:

OP_FREQ_XXXXXX_YYYYYY_INST_PLVL_ID.csv

where,

- OP is a two-digit code for the observed property;
- FREQ is a three-digit code for the sampling frequency;
- XXXXXX is a six-digit X lambert coordinate;
- YYYYYY is a six-digit Y lambert coordinate;
- INST is a three-digit code for the data provider;
- PLVL is a two-digit code for the product level;

The metadatabase contains 19 different observed properties aligned with the same unit reference, as shown in Table II.1.

Table II.1: MIGRADAKH metadatabase observed properties

Code	Description	Unit
SD	Stream discharge, cubic meters per second	m^3/sec
SL	Stream water level above DNG datum	m
AT	Temperature, air, degrees Celsius	deg C
SR	Total solar radiation, watts per square meter	W/m^2
WS	Wind speed in meter per second	m/sec
RH	Relative humidity, percent	%
GL	Groundwater level elevation above DNG datum, corrected for barometric pressure, meter	m
RF	Precipitation in millimeter	mm
RG	Relative gravity, nanometer per square second	nm/s^2
WT	Temperature, water, degrees Celsius	deg C
AP	Atmospheric pressure, hecto pascals	hPa
WP	Water pressure, hecto pascals	hPa
ST	Temperature, soil surface, celcius degree	deg C
DT	Temperature, deep soil, degrees Celcius	deg C
ET	Potential evapotranspiration, millimeter	mm
DD	Cave drip water discharge, Liter per hours	L/h
VW	Volumetric water content, L^3 of water per L^3 of soil	-
ER	Electrical resistivity, ohm meter	Ohm.m
WC	water electrical conductivity, millisiemens per centimeter	mS/cm

The sampling frequency is always a fixed frequency following Python Pandas convention for frequency string (https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html). If the data sent by the provider is monitored at unequal frequencies, the sampling frequency is either the highest sampling frequency or a lower frequency if the data was resampled.

Data provider codes are DG2 (SPW-Direction générale opérationnelle 2), DG3, (SPW- Direction général opérationnelle 3), ORB (Observatoire Royale de Belgique), PMB (PAMESEB asbl), UMO (Université de Mons) and UNA (Université de Namur). Table II.2 provides a summary of the database/metadatabase content by providers and observed properties. The database and metadatabase contents describe 123 time-series and 49 monitoring features. Regarding UMO, only the mean time series of the Electrical Resistivity Tomography data was encoded, but the full dataset is available from online repositories (see Watlet et al., 2018a). Regarding ORB, the RG data were provided by Arnaud Watlet. These are corrected with the pressure effect with an admittance of -3.3 $nm.s^{-2}.hPa^{-1}$. The product level code is referenced in the next table:

Table II.2: Summary of MIGRADAKH database per observed properties and providers

	DG2	DG3	ORB	PMB	UMO	UNA
AP						1
AT				1		1
DD					4	
DT				1		
ER					1	
ET				1		
GL						24
RF	2		1	1		
RG			2			
RH				1		
SD	2	6				4
SL	2	6				
SR				1		
ST				1		
VW					3	
WC						13
WP						13
WS				1		
WT		6				24

Table II.3: MIGRADAKH database product-level code

Code	Description
A0	Measured and automatically stored
A1	Measured and automatically stored after validation routine
AX	A product, unknown level
B0	Validated original measurements
B1	Validated resampled measurements with no attempts to fill missing values using an interpolation method
B2	Validated resampled measurements containing interpolated missing values
BX	B product, unknown level
C0	Uncalibrated simulated data
C1	Calibrated simulated data
C2	Inversed modeled data
C0	C product, unknown level

Appendix III. Chapter II List of Indexed Definitions

Table III.1: Causality-related definitions indexed in Chapter 2.

Index	Definition	Reference
d1	Something or someone that produces an effect, result, or conditions (*cause-effect relationships*, e.g., heavy rainfall causes floods).	Merriam-Webster's Learner's Dictionary
d2	A reason for doing or feeling something (*perceptual causes*, e.g., floods prevent someone from building in risky areas).	Merriam-Webster's Learner's Dictionary
d3	Something (such as an organization, belief, idea, or goal) that a group or people support or fight for (*teleological causes*, e.g., flood risks should be reduced in the future).	Merriam-Webster's Learner's Dictionary
d4	(*knowledge* is) perception.	Plato's *Theaetetus*; Chappell, 2019.
d5	(*knowledge* is) true belief.	Plato's *Theaetetus*; Chappell, 2019.
d6	(*knowledge* is) justified true belief.	Plato's *Theaetetus*; Chappell, 2019.
d7	*Material cause*: "that out of which", e.g., the bronze of a statue.	Aristotle's *Physics*; Falcon, 2019; Shields, 2008.
d8	*Formal cause*: "the form", "the account of what-is-to-be", e.g., the shape of the statue.	Aristotle's *Physics*; Falcon, 2019; Shields, 2008.
d9	*Efficient cause*: "the primary source of the change or rest", e.g., the sculptor, the art of bronze-casting the statue.	Aristotle's *Physics*; Falcon, 2019; Shields, 2008.
d10	*Final cause*: "the end, that for the sake of which a thing is done", e.g., art, representing someone, for a buyer.	Aristotle's *Physics*; Falcon, 2019; Shields, 2008.
d11	*Abstraction*: An abstract mental object on which we reason, e.g., a reservoir in a model, the concept of rain, an equation.	Personal definition
d12	*Deduction*: inference in which a new abstraction (thesis, d11) about particulars follows necessarily from prior general or universal abstractions (hypotheses), e.g., since rain falls, the ground will be wet.	Personal definition

Table III.1 (next-1): Causality-related definitions indexed in Chapter 2.

Index	Definition	Reference
d13	*Induction*: inference of a generalized abstraction (d11) from the sensory observation of particular instances in the real world, e.g., rain is always followed by the ground being wet; the fact that rain precipitates; what rain itself is in general, an empirical equation.	Personal definition
d14	*Abduction*: inference to the most likely general and universal premises associated with a conclusion generalized from particular instances, e.g., the ground is wet, it must have rained.	Personal definition; Peirce, 1960.
d15	*Imagination*: the mental ability to create abstractions not exclusively based on sensed particular instances, e.g., an equation; a virtual abstract reservoir; an elephant watering the ground.	Personal definition
d16	*Actuality* (*entelecheia* or *energeia*): what is capable of being seen, the account of what is seen, e.g., the statue of bronze, the artisan crafting the statue, data;	Aristotle's *Physics* and *Metaphysics*; Cohen, 2016; Marmodoro, 2018.
d17	*Potentiality* (*dunamis*) or *dispositionality*: what is capable of being built, the capacity to be in a different state, the account of what-could-be, e.g., a sword of bronze, forecasts.	Aristotle's *Physics* and *Metaphysics*; Cohen, 2016; Marmodoro, 2018.
d18	*Principle of parsimony*, or *Ockham's razor*: one should not multiply reasons without necessity.	Baker, 2016.
d19	*Principle of sufficient reason*: everything has a cause.	Leibniz; Melamed and Lin, 2018.
d20	*Necessary cause*: a cause/reason/condition that is always involved in the realization/observation of an event/state;	Melamed and Lin, 2018.
d21	*Sufficient cause*: a cause/reason/condition or a set of it that always implies the realization/observation of an event/state.	Melamed and Lin, 2018.
d22	*Principle of priority*: the cause occurs before the effect, e.g., rainfall occurs before flood peaks.	Hume (1738, 1748).
d23	*Principle of contiguity*: the cause and effect are contiguous in time and space, e.g., the observed rainfall and flood peaks are closely related in space and time.	Hume (1738, 1748).

Table III.1 (next-2): Causality-related definitions indexed in Chapter 2.

Index	Definition	Reference
d24	*Principle of constant conjunction*: the occurrence of the cause systematically implies the occurrence of the effect [under the same conditions], e.g., similar rainfall always implies the observation of similar flow peaks under similar conditions.	Hume (1738, 1748).
d25	*Necessary connection*: the additional principle that is necessary to avoid being deceived by the first three, e.g., we cannot know ultimately if rain causes flood peaks, but we can speculate infinitely deeper and deeper on overland flow, subsurface flow, hydrological connectivity, etc.	Hume (1738, 1748).
d26	*Principle of the common cause*: constant conjunction (d24) is either the product of causal interrelation or the result of shared driving variables.	Reichenbach , 1956.
d27	*Manipulable cause*: handling devices to manipulate the effects.	Woodward, 2016.
d28	*Dispositional cause*: a cause that is attributed internally to one's personal trait (d2, >< d29), e.g., the manager is incompetent.	Heider, 1958; Kelley, 1973; Weiner et al., 1987.
d29	*Situational cause*: a cause that is attributed externally to the environment (d2, >< d28), e.g., the manager was overwhelmed; It was an extreme rainfall.	Heider, 1958; Kelley, 1973; Weiner et al., 1987.
d30	*Stable cause*: a cause that is perceived as temporally persistent (d2, >< d31), e.g., the manager incompetence (if deemed persistent); heavy rainfall and floods always happen.	Heider, 1958; Kelley, 1973; Weiner et al., 1987.
d31	*Unstable cause*: a cause that is perceived as being temporary or rare (d2, ><d30), e.g., the manager was inexperienced; It has rained a lot the past few weeks.	Heider, 1958; Kelley, 1973; Weiner et al., 1987.
d32	*Controllable cause*: a cause that is perceived as being manipulable (d2, d27, >< d33), e.g., the manager can learn, rain or urbanism (if trust in management).	Heider, 1958; Kelley, 1973; Weiner et al., 1987.
d33	*Uncontrollable cause*: a cause that is perceived as being not manipulable (d2, >< d27, >< d32), e.g., incompetency (if deemed persistent), rain (without trust in management).	Heider, 1958; Kelley, 1973; Weiner et al., 1987.

Table III.1 (next-3): Causality-related definitions indexed in Chapter 2.

Index	Definition	Reference
d34	*Causal frame*: the way (mostly unconscious) in which the individual mind conceptualizes a stable causal model of their perceived reality (d2), e.g., how floods work.	Chong and Druckman, 2007; Lakoff, 2010; Tversky and Kahneman, 1981.
d35	*Values*: guiding principles in people's life (d2,d3), e.g., individualism, altruism, ecologism embodied in opinions about flood management.	Stern, 2000; Stern et al., 1999; Roobavannan et al., 2018.
d36	*Beliefs*: beliefs about what is true or not (d2, d5), e.g., "urbanism causes floods."	Stern, 2000; Stern et al., 1999; Roobavannan et al., 2018.
d37	*Norms*: rules, either formal or informal, that prescribes people's behavior (d2, d3), e.g., the flood directive, cultural habits related to the water system.	Stern, 2000; Stern et al., 1999; Roobavannan et al., 2018.
d38	*Static teleology*: a static arrangement that seems to be useful for a certain "purpose", e.g., optimality principle in ecohydrological distribution of plant species in watersheds (Eagleson, 2002); slopes directed to the outlet.	von Bertalanffy, 1968.
d39	*Asymptotic teleology*: asymptotic behavior that attains a time-independent condition, e.g., a storm basin depletion.	von Bertalanffy, 1968.
d40	*Emergent teleology*: a purpose that arises from/in the organized structure, e.g., watershed's function as a support for life, ecosystems, and human systems (Kumar, 2007); physical habitat, food supply (Sivapalan, 2006); water partition, storage, release (Wagener et al., 2007).	von Bertalanffy, 1968.
d41	*Equifinal teleology*: a final state that can be reached from different initial conditions, e.g., Beven's equifinality of model space (Beven, 2006a); a steady baseflow; equifinal flow paths to the outlet.	von Bertalanffy, 1968.
d42	*True teleology*: a true anticipated goal, e.g., water laws, directives, or acts; people's behaviors.	von Bertalanffy, 1968.

Table III.1 (next-4): Causality-related definitions indexed in Chapter 2.

Index	Definition	Reference
d43	*Causal explanation*: a stable agreement, perception or belief emerging from a collective thinking constrained by logic in interaction with the real world [d4, d5, d6, d25, d28, d30, d34, d36, d38-42] explaining within the collective context and for a specific purpose [d2, d3, d10, d29, d35, d37, d39, d42] the what-is-to-be [d8, d11, d16, d38-d42] through mechanisms [d1, d9, d19, d24] linked by inference to other elements [d7, d12-15], either in space [d23], in time [d22], or both and possibly at other scales, in an intelligible but sufficient way [d18, d21, d26] and by virtue of a potential future practical application [d2, d3, d10, d17, d27, d32, d41] enabling either testability, technological progress or successful control while operating and intervening on the real world.	Personal definition

Appendix IV. Cross-Predictive Patterns in Hydrograph Separation

The Eckhardt filter is a popular linear method for hydrograph separation into quickflow and baseflow [1]. For a streamflow time series Q_t, its baseflow component B_t is obtained using a recursive filter depending on both Q_t and B_{t-1}:

$$B_t = \frac{(1-BFI_{max})\alpha B_{t-1}+(1-\alpha)BFI_{max}Q_t}{1-\alpha BFI_{max}}, \text{ subject to } B_t \leq Q_t$$

where α is the recession constant for a linear recession following $Q_t = Q_0\alpha^t$, and BFI_{max} is the maximum baseflow index, i.e., the long-term mean proportion of baseflow in the streamflow. A quickflow component A_t is simply obtained considering that $A_t = Q_t - B_t$. The α parameter can be estimated from recession analysis. However, the BFI_{max} is harder to estimate based on hydrograph time-series analysis. In practice, rough guidelines exist based on dominant catchment geology, rough estimates can be obtained from percentile statistic, or one has to get involved in time-consuming and sporadic tracing tests. The effect of these two parameters on the baseflow separation is illustrated in Figure IV.1 (with BFI_{max} denoted as β).

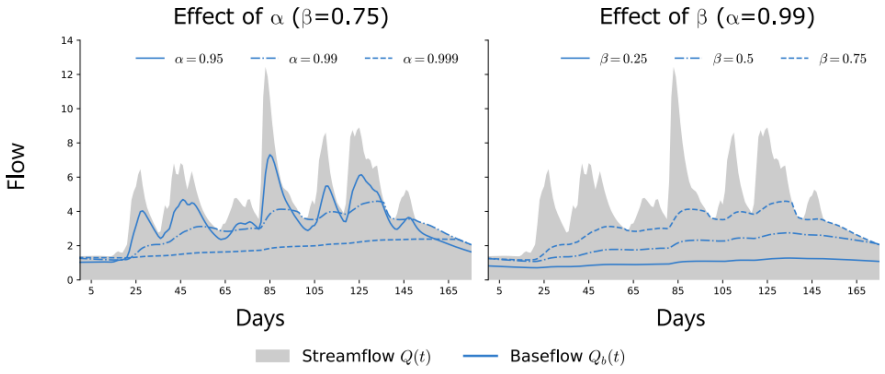

Figure IV.1: Illustrated effects of the recession constant α and the maximum baseflow index β on the output of hydrograph separation using the Eckhardt filter.

Before focusing exclusively on recession analysis, the initial work related to Chapter 4 also investigated hydrograph separation with the Eckhardt filter. Using synthetic streamflow series, the analysis revealed interesting cross-

predictive patterns between A_t and B_t that depends on the BFI_{max} parameter and could therefore be used for the parametrization of the Eckhardt filter (Figure IV.2.A). Instantaneous cross-prediction were made with the bivariate extension of the EDM-Simplex algorithm, also known as Convergent Cross Mapping [2], from reconstructed Eckhardt quickflow A_t to its corresponding Eckhardt baseflow series B_t with an embedding dimension of 3. Five virtual streamflow series of constant α and known BFI_{max} ranging between 0.32 to 0.82 were decomposed into 30 baseflow series using linearly increasing BFI_{max} from 0.1 to 0.97. The results of the virtual experiment (Figure IV.2.A) exhibit a clearly discernible pattern: whenever the Eckhardt BFI_{max} parameter is close to the value obtained from the model generating the synthetic series, prediction skills from quickflow to baseflow tends to be minimum. On the real Lhomme dataset (section S1, S2, S3), a initial decay of forecasting skills is observed up to a point that could be suggested as a proper BFI_{max} (Figure IV.2.B), considering the patterns of Figure IV.2.A. However, no strong rebound of predictive skill is observed, as in Figure IV.2.A, and the patterns in Figure IV.2.B are rather assimilable to a kneel.

Although these patterns are interesting and, I believe, correlate with the actual baseflow index, the analysis was not included in the published version of the paper or chapter 4 because: (1) considering both recession analysis and hydrograph separation with virtual and real experiments made the paper cumbersome, (2) the idea of combining a nonlinear method for parameterizing a linear filter is somewhat far-fetched, and (3) the patterns did not find enough explanatory elements to enlighten their form.

[1] Eckhardt, K. « How to Construct Recursive Digital Filters for Baseflow Separation ». Hydrological Processes 19, no 2 (2005): 507-15. https://doi.org/10.1002/hyp.5675.

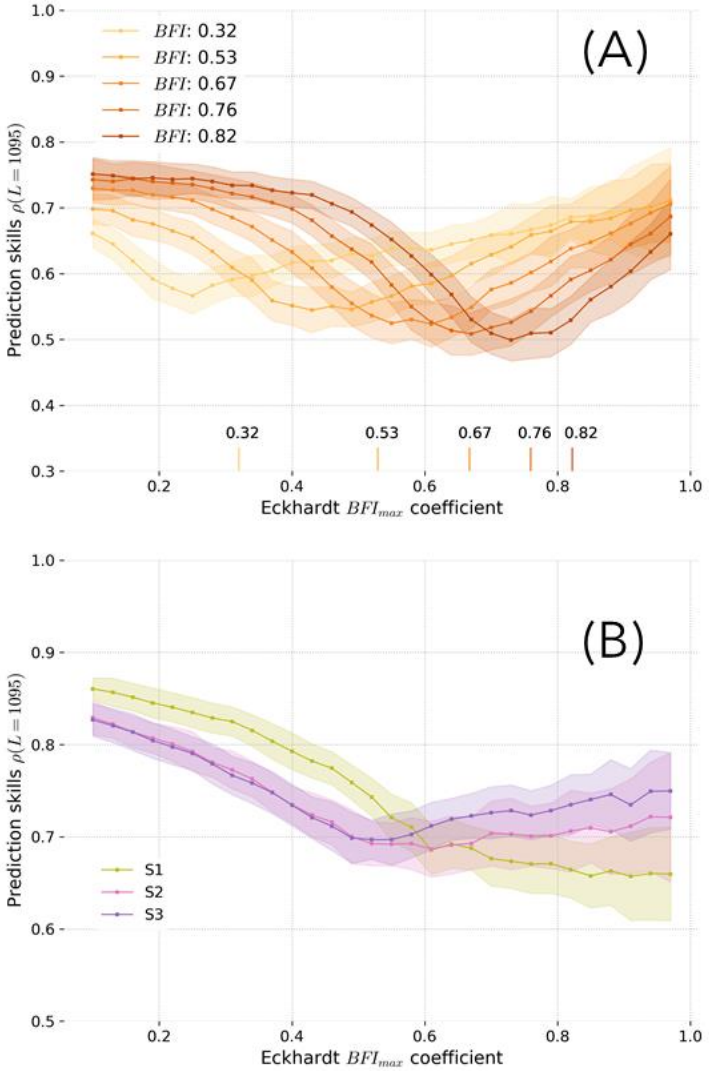

Figure IV.2: Eckhardt baseflow prediction skills from Eckhardt quickflow applied to (A) synthetic streamflow - Each color corresponds to a virtual series defined by its known baseflow index (BFI). Actual BFI from the model are reported on the X-axis, allowing to relate the prediction skills pattern across the range of BFI_{max} to the original value. -; (B) to the Lhomme dataset. Each color corresponds a Lhomme river gauging station ordered from upstream to downstream: S1(lime), S2 (pink), S3 (violet). For each subplot, colored band represents 3 standard deviations of CCM prediction skills.

Appendix V. Example of Empirical Recession Extraction

Figure V.1 illustrates how recession points can be identified based on EDM-Simplex self-predictions for station S3. The hyper-parametrization of EDM-Simplex is different from the final one presented in section 4.2.3. For instance, Figure V.1 considers a prediction horizon tp of 1 day, instead of 0 days, explaining the higher error threshold (around -6 instead of -10). The underlying logic and philosophy of extraction remain the same: recession points are the most deterministic and most self-predictive points in the hydrograph. Therefore, the method extract recession points as hydrograph points with the best median self-predictive skills, i.e., those lying on the identity line in Figure V.1.a. In practice, the point are extracted using a threshold value on the error function $\log(\varepsilon^2/Q)$ where ε are the residuals between the observed Q and the median EDM-Simplex prediction. Dividing by Q allows to be more tolerant for high Q values, that are early recession points presenting a more uncertain dynamic. The threshold is set at -6 six such that recession points correspond to an unbiased model, meaning that the Esperance of residuals $\hat{E}[\varepsilon]$ should be around zero (Figure V.1.b). To avoid remaining anomalies or isolated recession points, a post-processing criterion is used to keep recession segments of minimum 2 days. The results for the first two years are shown in Figure V.1.c.

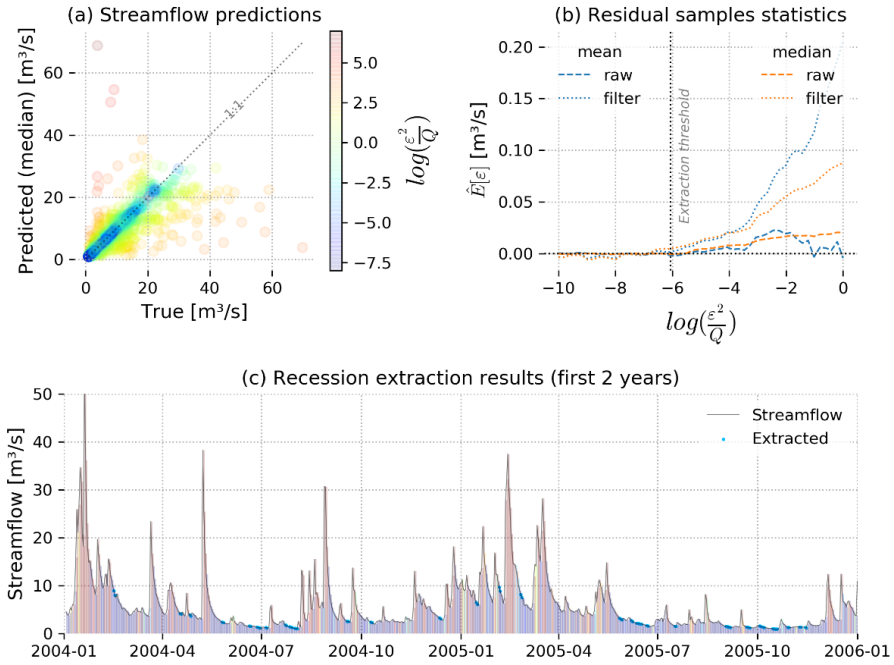

Figure V.1: Empirical recession extraction for station S3 using EDM-Simplex: (a) Scatterplot of the median EDM-Simplex prediction at $t + 1$ applied on the full S3 time series. Errors are represented using the logarithm of the squared residuals ε^2 divided by Q. Deterministic recession points are expected to fit on the identity line with low errors (blue points); (b) Error threshold versus the mean and median of residuals sampled below the threshold. The final threshold is selected so that the EDM model for recession is unbiased both on raw data (raw) and on segments of a minimum length of 2 days (filter) ; (c) Results of the extraction on the first two years of S3 (blue points) for the segments of a minimal length of at least 2 days. The color under the streamflow curve map to the color bar of (a) and represents the inherent uncertainties associated with recession.

Appendix VI. Chapter 6 Supplementary Figures

A. Virtual experiment

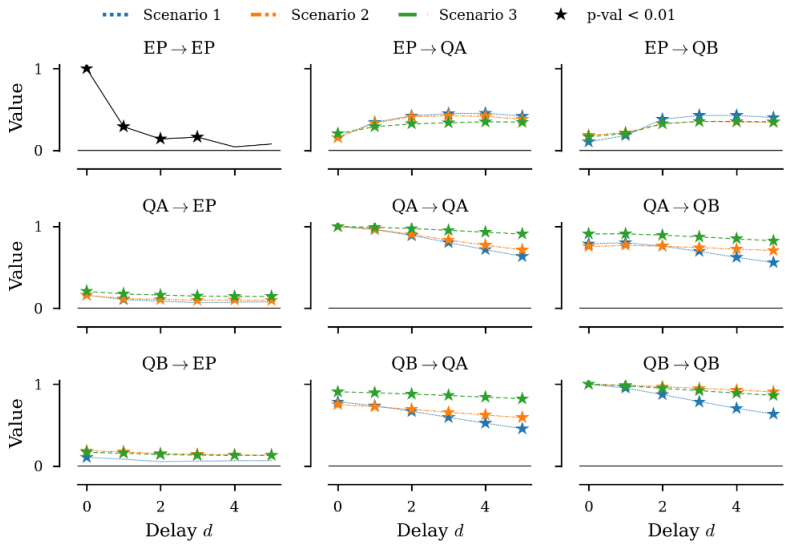

Figure VI.1: Time-dependencies for the CCF method on raw data with a noise level factor of 0.15

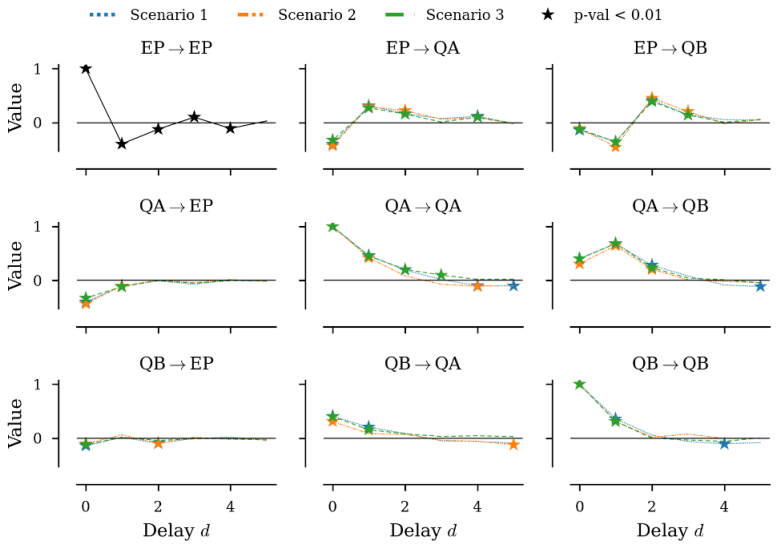

Figure VI.2: Time-dependencies for the CCF method on differenced data with a noise level factor of 0.15

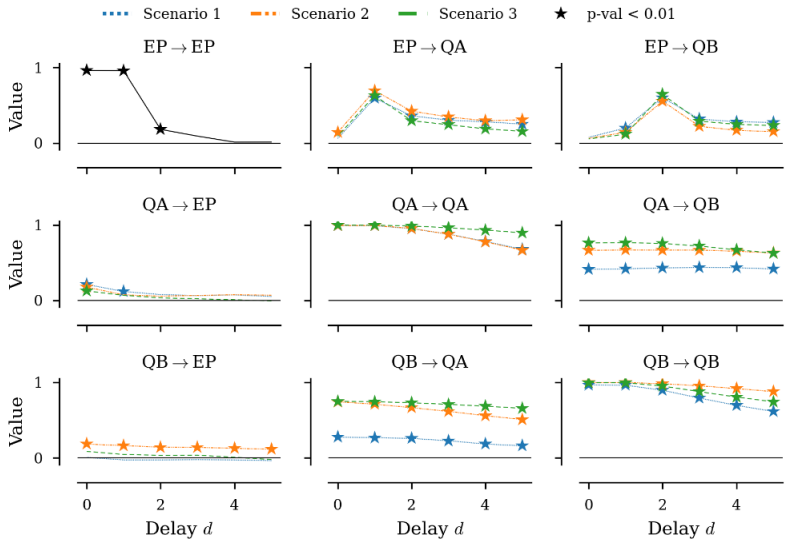

Figure VI.3: Time-dependencies for the CCM method on raw data with a noise level factor of 0.15. CCM was applied with an embedding dimension of 2, an embedding delay of 1, and a library length of 200.

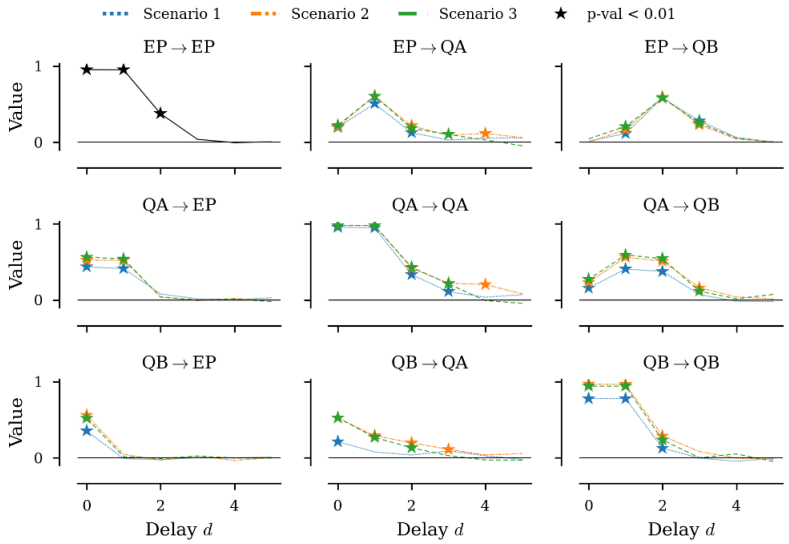

Figure VI.4: Time-dependencies for the CCM method on differenced data with a noise level factor of 0.15. CCM was applied with an embedding dimension of 2, an embedding delay of 1, and a library length of 200.

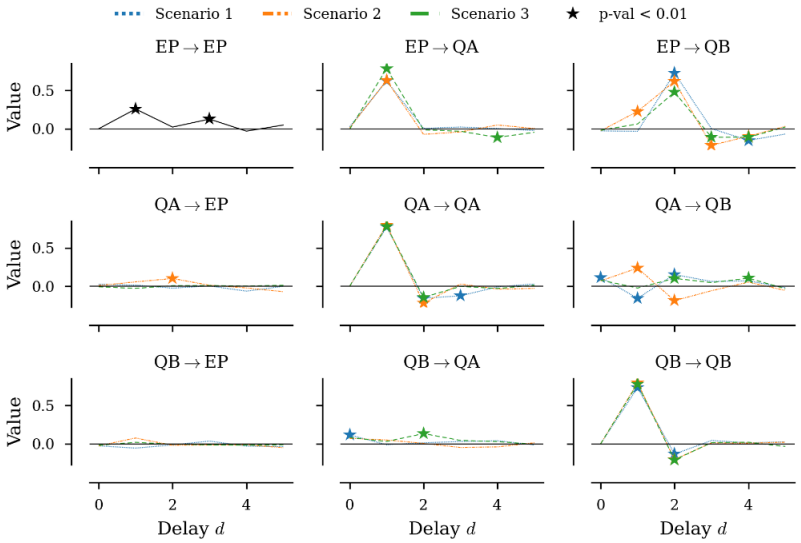

Figure VI.5: Time-dependencies for the ParCorr method on raw data with a noise level factor of 0.15.

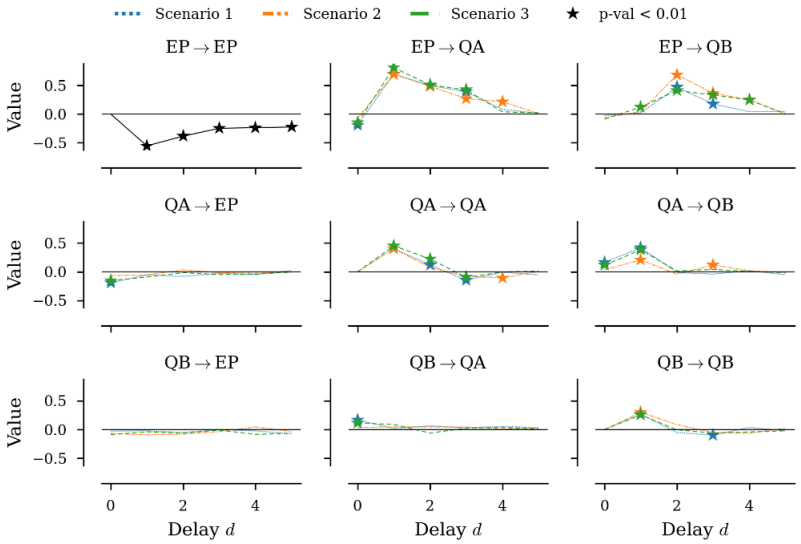

Figure VI.6: Time-dependencies for the ParCorr method on differenced data with a noise level factor of 0.15.

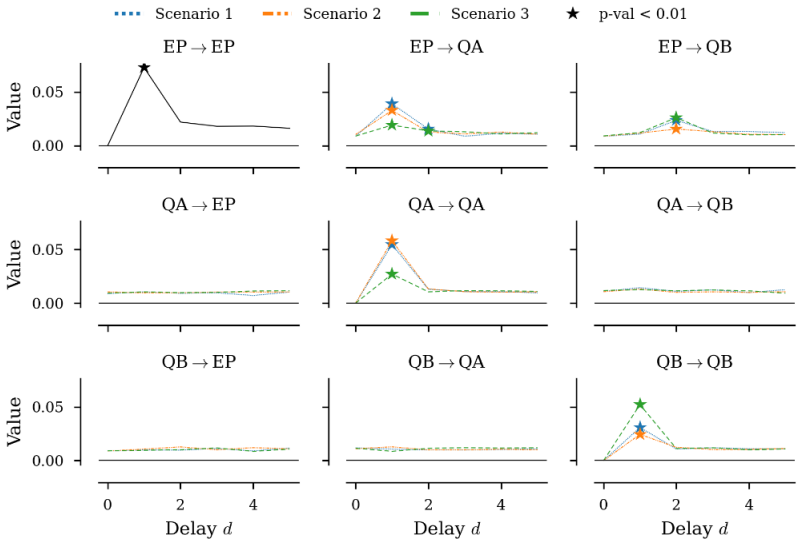

Figure VI.7: Time-dependencies for the CMI method on raw data with a noise level factor of 0.15 and a k_{CMI} of 20.

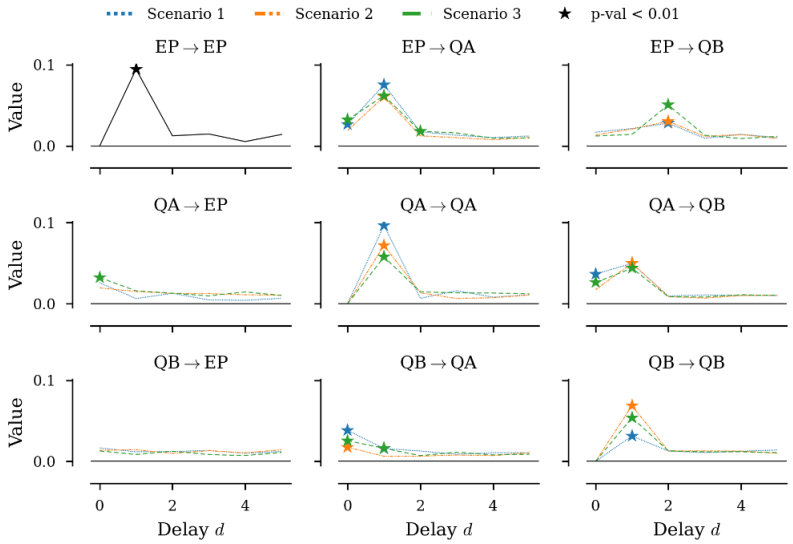

Figure VI.8: Time-dependencies for the CMI method on differenced data with a noise level factor of 0.15 and a k_{CMI} of 20.

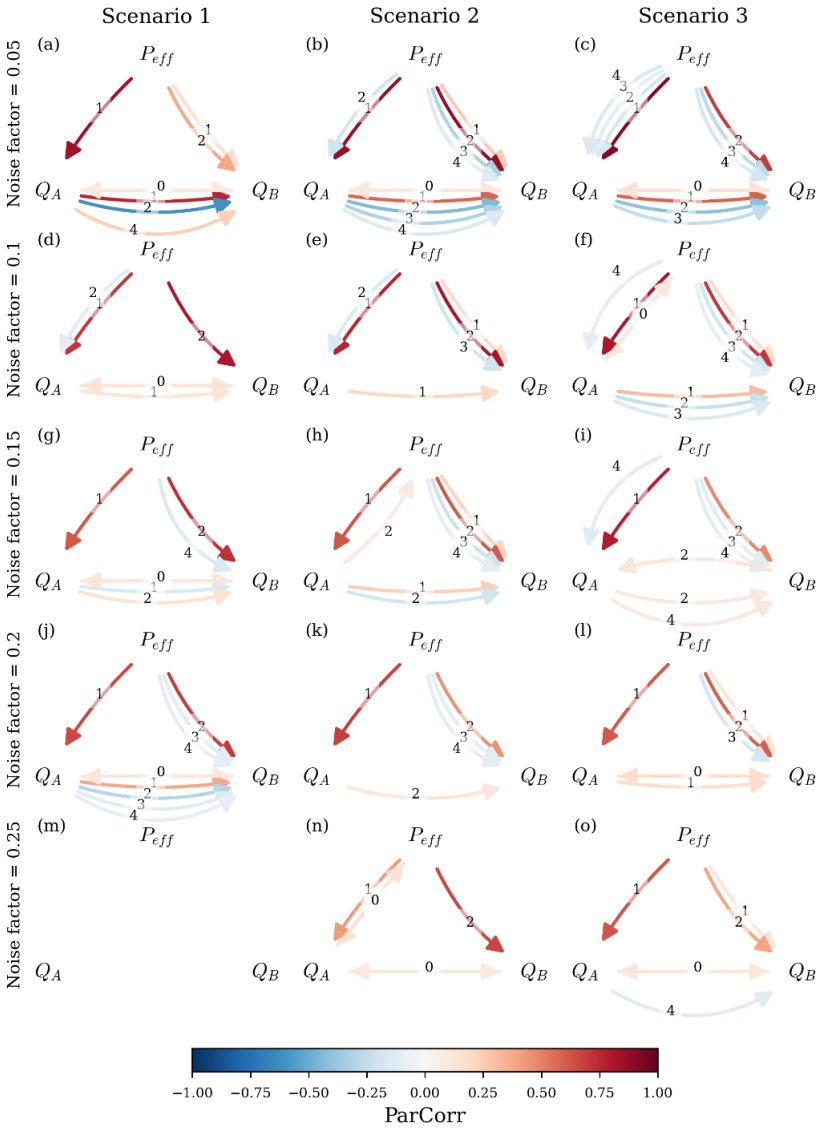

Figure VI.9: Influence of noise level on ParCorr resulting DAG of cross-dependencies. The individual plots (a to o) indicate the significant causal relationship (p-val < 0.01) using directional arrows whose color corresponds to the value of the partial correlation. Causal delays are annotated next to each arrow. ParCorr is applied to each raw dataset corresponding to the scenarios of Table 6-2 (columns) for different multiplicative noise factors (rows).

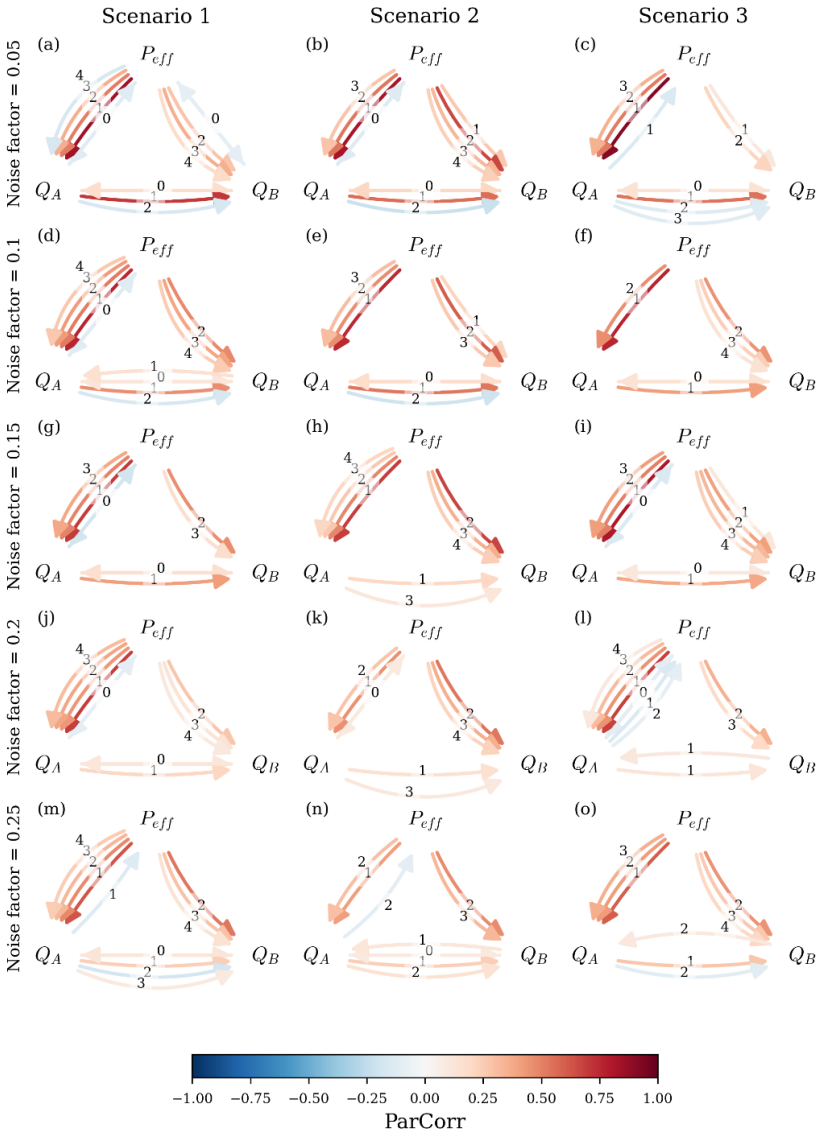

Figure VI.10: Directed Acyclic Graph of ParCorr cross-dependencies. The individual plots (a to o) indicate the significant causal relationship (p-val < 0.01) using directional arrows whose color corresponds to the value of the partial correlation. Causal delays are annotated next to each arrow. ParCorr is applied to each differenced dataset corresponding to the scenarios of Table 6-2 (columns) for different multiplicative noise factors (rows).

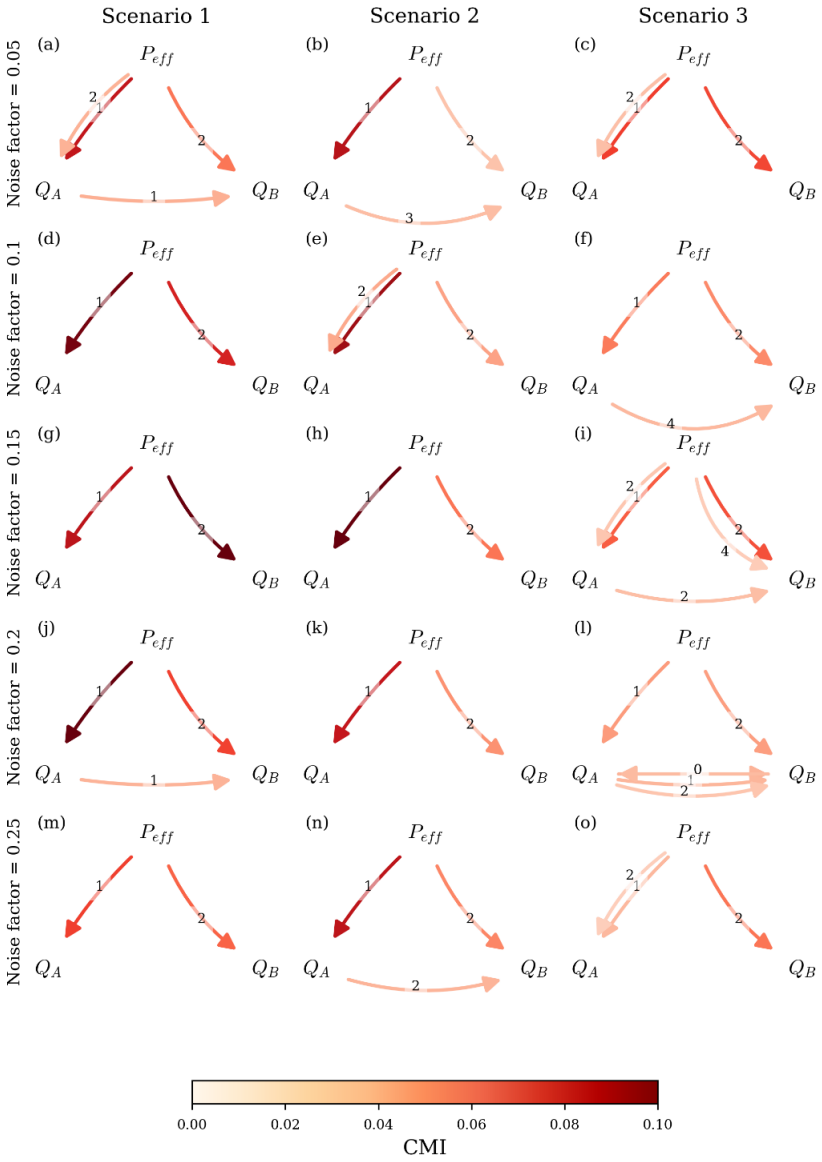

Figure VI.11: Directed Acyclic Graph of CMI cross-dependencies with a k_{CMI} parameter equal to 5. The individual plots (a to o) indicate the significant causal relationship (p-val < 0.01) using directional arrows whose color corresponds to the value of the partial correlation. Causal delays are annotated next to each arrow. The CMI test is applied to each raw dataset corresponding to the scenarios of Table 6-2 (columns) for different multiplicative noise factors (rows).

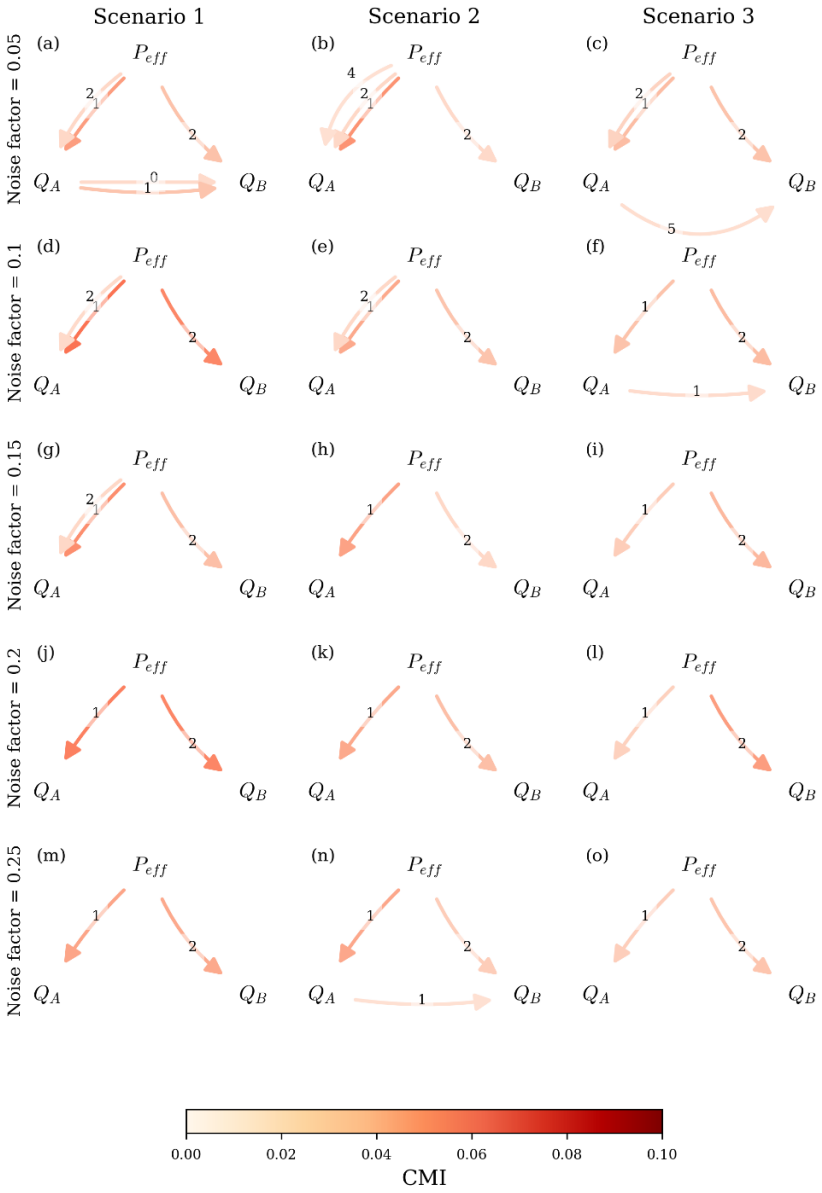

Figure VI.12: Directed Acyclic Graph of CMI cross-dependencies with a k_{CMI} parameter equal to 20. The individual plots (a to o) indicate the significant causal relationship (p-val < 0.01) using directional arrows whose color corresponds to the value of the partial correlation. Causal delays are annotated next to each arrow. The CMI test is applied to each raw dataset corresponding to the scenarios of Table 6-2 (columns) for different multiplicative noise factors (rows).

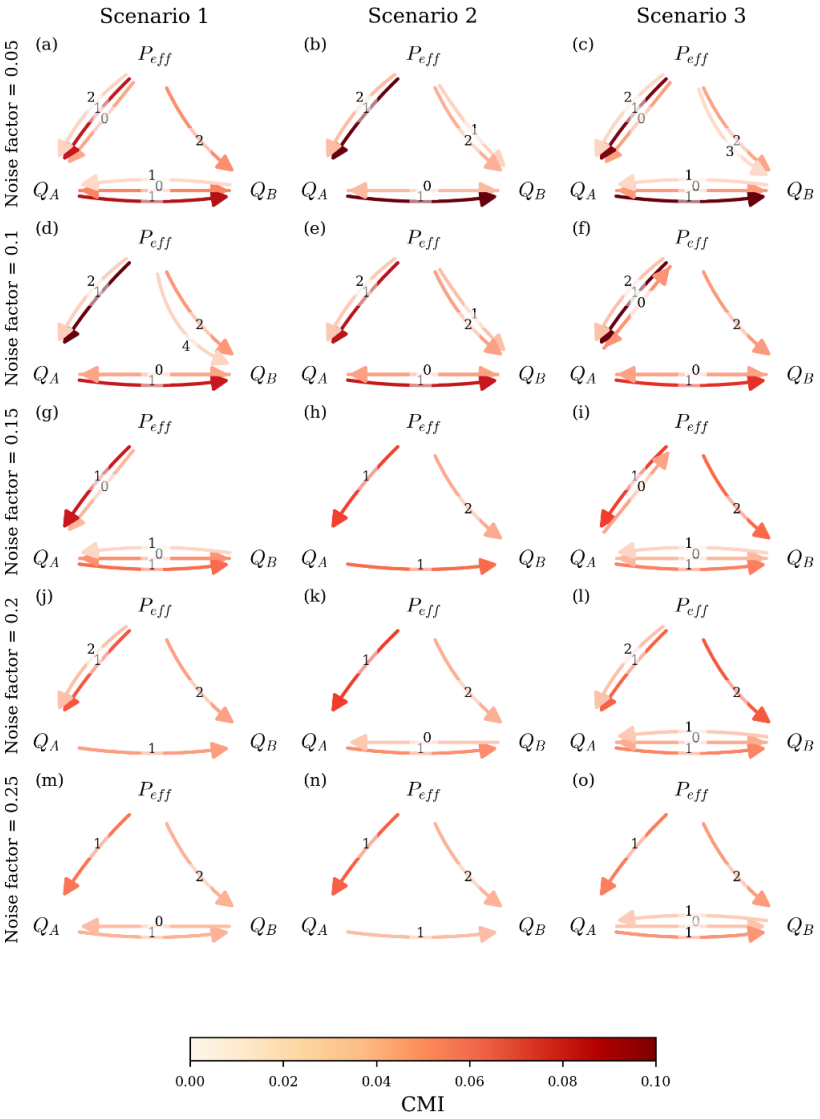

Figure VI.13: Directed Acyclic Graph of CMI cross-dependencies with a k_{CMI} parameter equal to 20. The individual plots (a to o) indicate the significant causal relationship (p-val < 0.01) using directional arrows whose color corresponds to the value of the partial correlation. Causal delays are annotated next to each arrow. The CMI test is applied to each differenced dataset corresponding to the scenarios of Table 6-2 (columns) for different multiplicative noise factors (rows).

B. Hydrological Connectivity in the Vadose Zone

Figure VI.14: CCF dependencies between all variables. All x-axis report the delay d and all y-axis the CCF Pearson correlation values. Red stars, orange squares, and green triangle reports respectively p-values below 0.001, 0.01, 0.05. Titles $X_t \rightarrow Y_t$ indicates which variable is the delayed potential driver (X_t) and the response variable (Y_t).

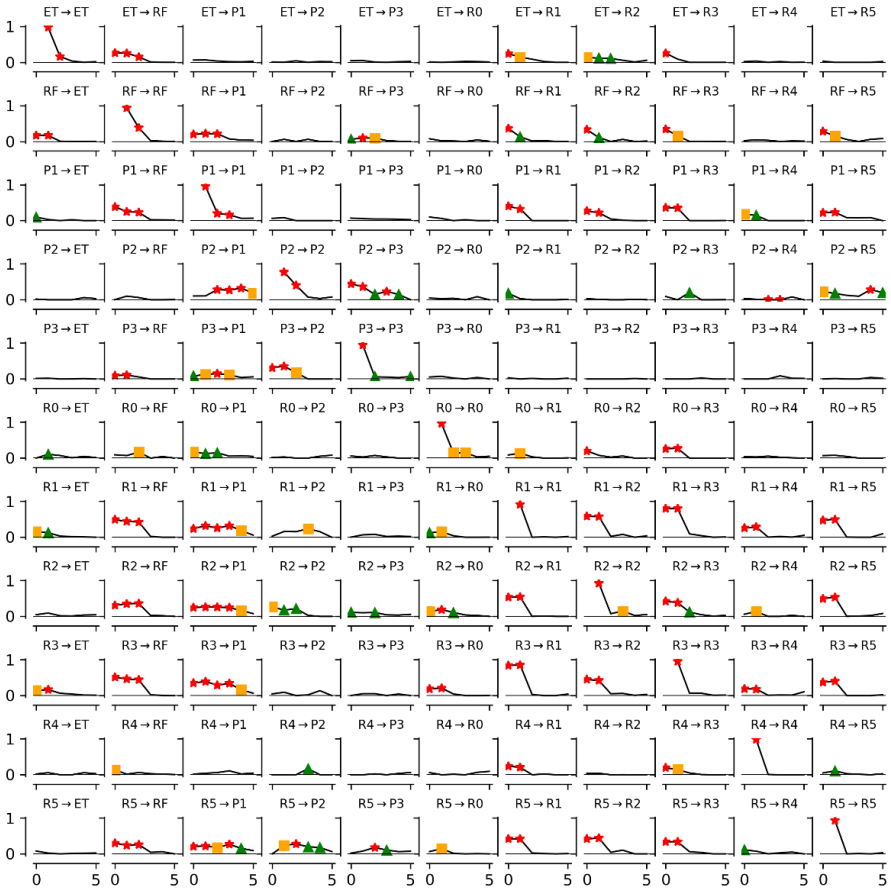

Figure VI.15: CCM dependencies between all variables. All x-axis report the delay d and all y-axis the CCM Pearson correlation values. Red stars, orange squares, and green triangle reports respectively p-values below 0.001, 0.01, 0.05. Titles $X_t \rightarrow Y_t$ indicates which variable is the delayed potential driver (X_t) and the response variable (Y_t). CCM was applied with an embedding dimension of 2, an embedding delay of 1, and a library length of 100.

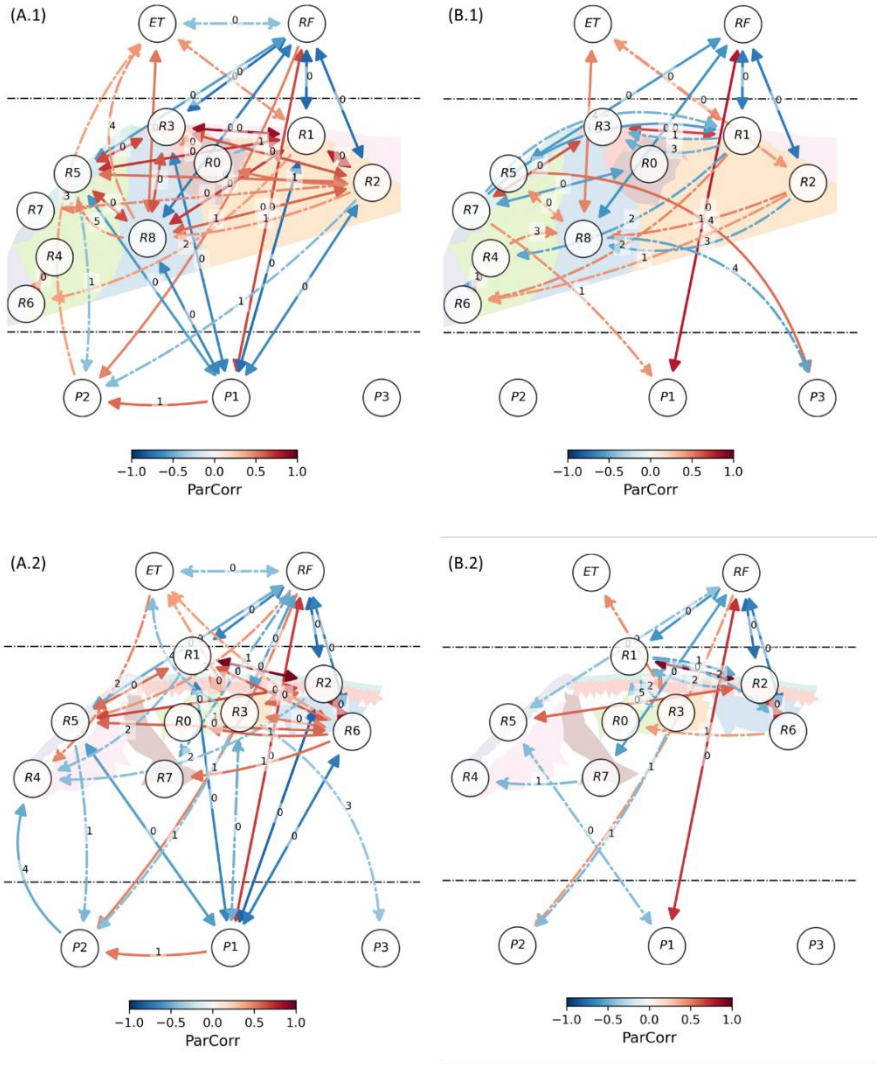

Figure VI.16: Graph of ParCorr cross-dependencies. (A.x) Unconstrained graph; and (B.x) constrained as follows: (1) ET and RF have no parents and are considered as exogeneous variables; (2) resistivity series cannot have a drip discharge series as parents; (3) a drip discharge series cannot have drip discharge series as parents except itself. Contemporaneous dependencies are represented by a bidirected straight arrow, delayed dependencies with curved arrows. All delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to ParCorr dependencies. Solid and dash-dotted arrows represent respectively significant dependencies with p-value < 0.001 and < 0.01. Variables are: evapotranspiration (ET), rainfall (RF), subsurface clustered resistivity time-series, and drip discharge data in the Rochefort Cave (P1 to P3). Resistivity clusters for A.1 and B.1 are those of Figure 5-6.n and their spatial extent of are displayed in the background using the same color palette. Resistivity clusters A.2 and B.2 are those from the expert classification of Figure 3-5 using the same color palette.

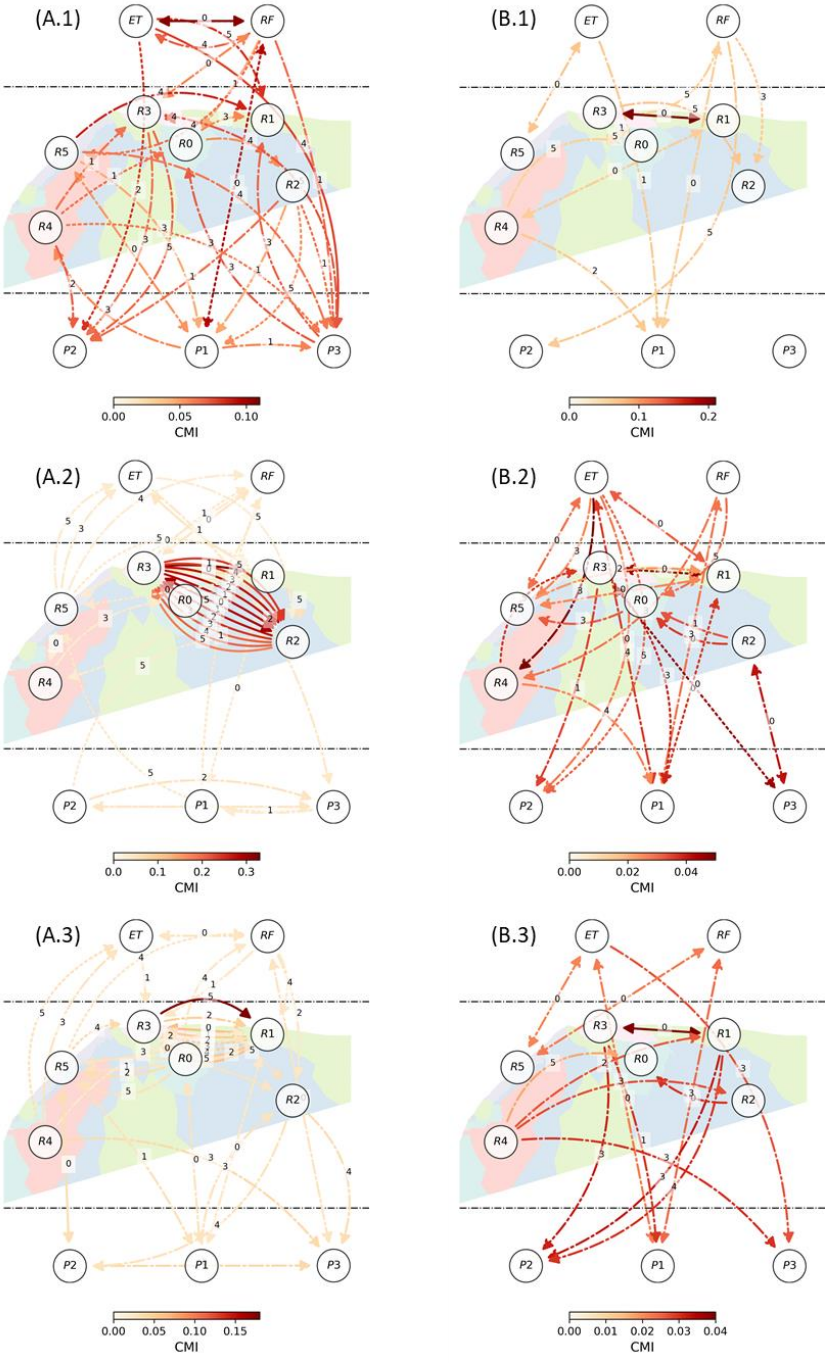

Figure VI.17: Graph of CMI cross-dependencies …

(A.x) Unconstrained graph; and (B.X) constrained as follows: (1) ET and RF have no parents and are just considered as exogeneous variables; (2) resistivity series cannot have a drip discharge series as parents; (3) a drip discharge series cannot have drip discharge series as parents except itself. Top row (A.1, B.1), middle row (A.2, B.2) and bottom row (A.3, B.3) uses respectively a k_{CMI} of 15, 30, and 35. Contemporaneous dependencies are represented by a bidirected straight arrow. Delayed dependencies are shown using directed curved arrows. All corresponding delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to CMI dependencies. Solid, dash-dotted, and dotted arrows represent respectively significant dependencies with p-value < 0.001, < 0.01, and < 0.05. Variables are: evapotranspiration (ET), rainfall (RF), subsurface clustered resistivity time-series (R0 to R5), and drip discharge data in the Rochefort Cave (P1 to P3). Resistivity clusters are those of Figure 5-6.f and their spatial extent of are displayed in the background using the same color palette.

C. Drivers of the Mass Balance

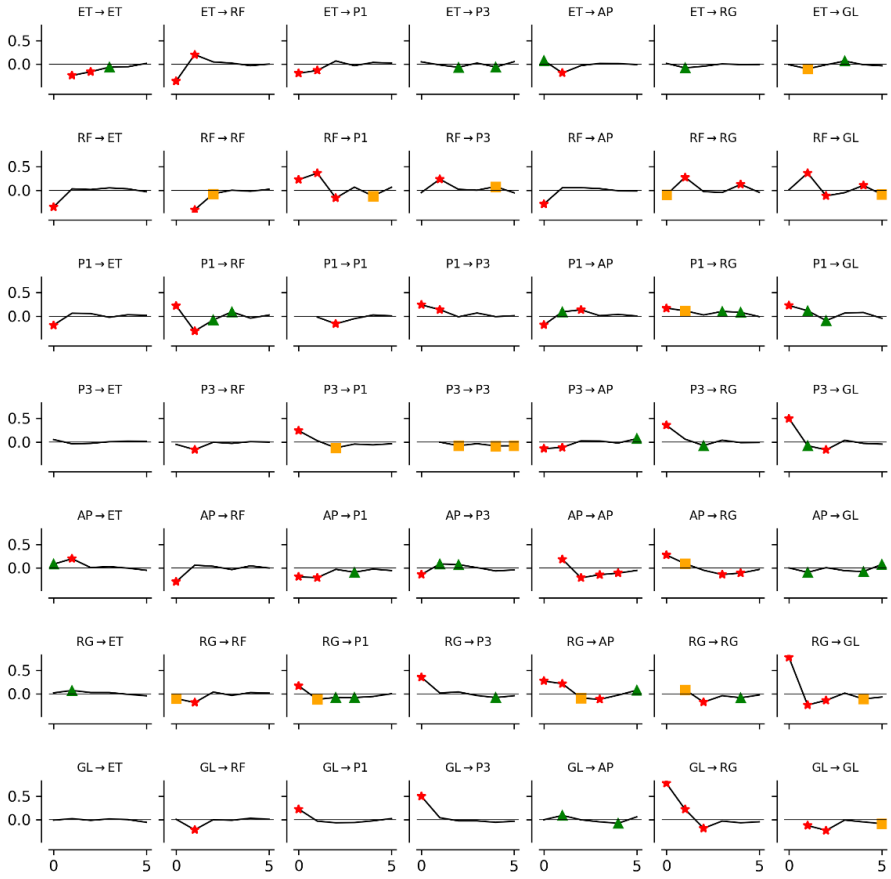

Figure VI.18:: CCF dependencies between all variables. All x-axis report the delay d and all y-axis the CCF Pearson correlation values. Red stars, orange squares, and green triangle reports respectively p-values below 0.001, 0.01, 0.05. Titles $X_t \rightarrow Y_t$ indicates which variable is the delayed potential driver (X_t) and the response variable (Y_t).

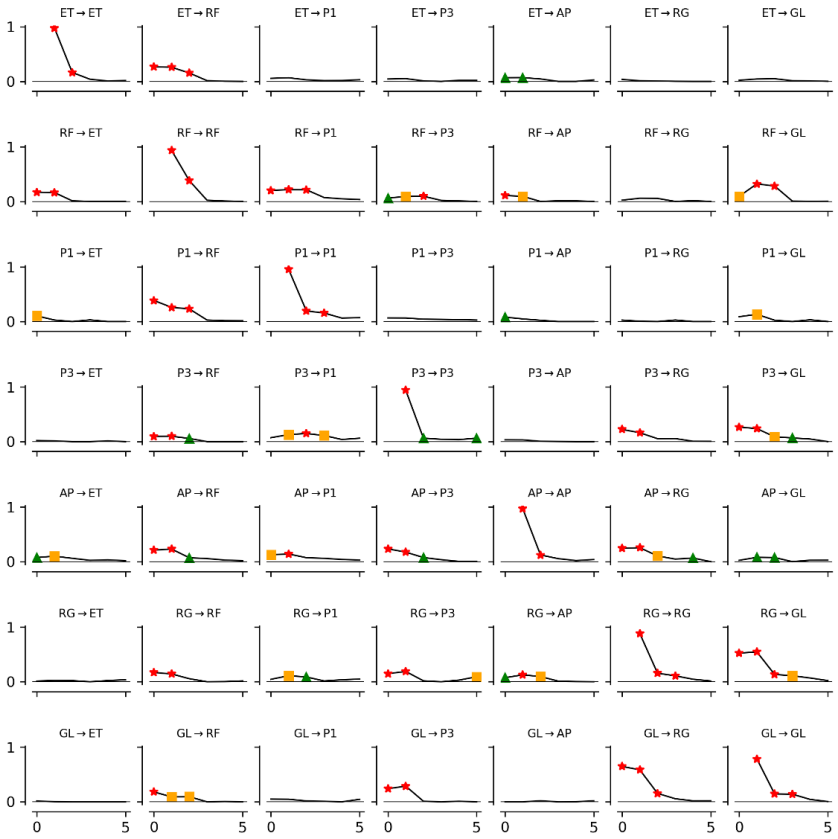

Figure VI.19: CCM dependencies between all variables. All x-axis report the delay d and all y-axis the CCM Pearson correlation values. Red stars, orange squares, and green triangle reports respectively p-values below 0.001, 0.01, 0.05. Titles $X_t \rightarrow Y_t$ indicates which variable is the delayed potential driver (X_t) and the response variable (Y_t). CCM was applied with an embedding dimension of 2, an embedding delay of 1, and a library length of 100.

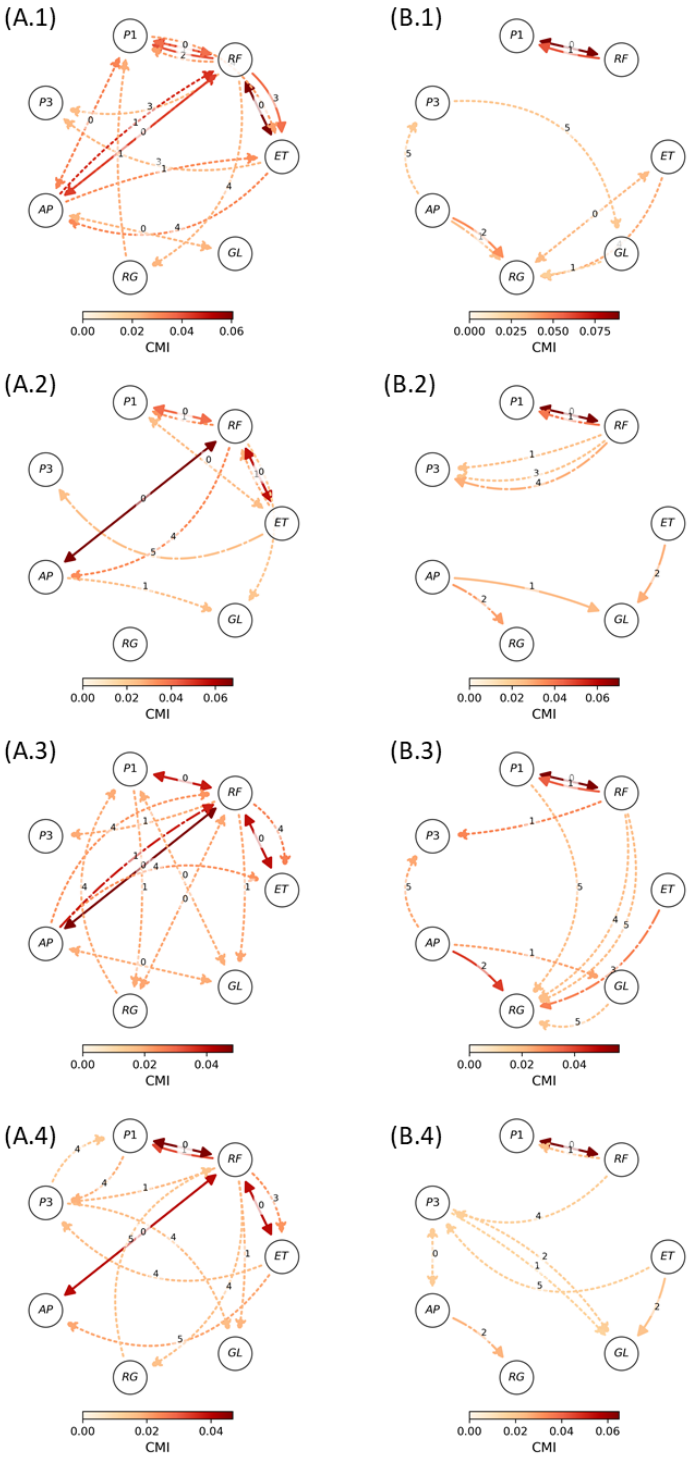

Figure VI.20: Graph of cross-dependencies …

(A.x) unconstrained graphs; (B.x) Constrained graphs CCM.; From the top row (A.1, B.1) to the bottom row (A.4, B.4), k_{CMI} varies as follows: 15, 20, 25, and 35. Variables are evapotranspiration (ET), rainfall (RF), groundwater level (GL), relative gravimetry (RG), atmospheric pressure (AP), and drip discharge data (P1, P3). Contemporaneous dependencies are represented by a bidirected straight arrow. Delayed dependencies are shown using directed curved arrows. All corresponding delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to the color bar scaling the time-dependencies. Solid, dash-dotted, and dotted arrows represent respectively significant dependencies with p-value < 0.001, < 0.01, and < 0.05.

Appendix VII. Pictures

Photo VII.1: The Vauclusian spring of Eprave (forefront) returning water to the Lhomme River (background). Source: KARAG (www.karag.be)

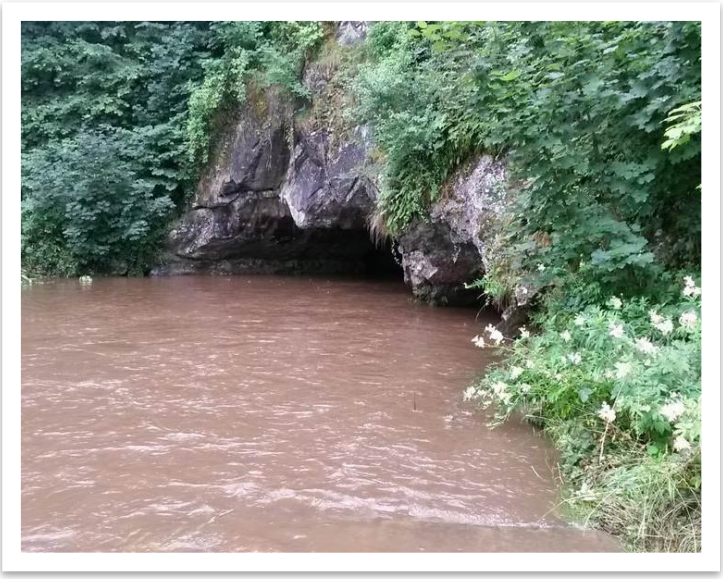

Photo VII.2: The Lhomme River (forefront) overflowing into the Nou Maulin cave (background) and flooding the cave system at Rochefort. Source: KARAG (www.karag.be)

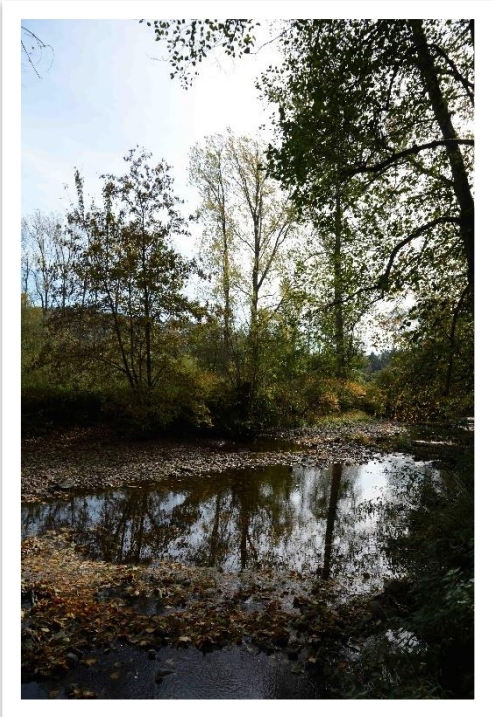

Photo VII.3: The Lhomme River, just upstream to the Vauclusian spring of Eprave, with almost no flow on the 12th of October, 2018, after a warm summer. Source: UNamur (Gaëtan Rochez, Amaël Poulain).

Photo VII.4: Shelter housing the relative superconducting gravimeter at the entrance of the Rochefort Cave Laboratory. Source: ROB (Michel Van Camp).

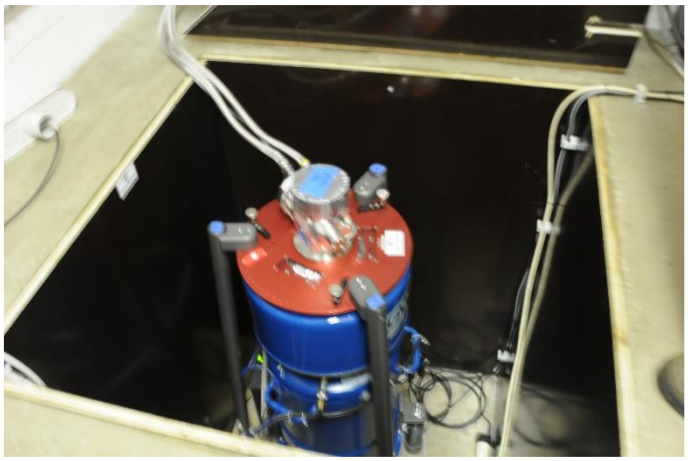

Photo VII.5: The relative superconducting gravimeter, with its shelter. Source: UCLouvain (Damien Delforge).

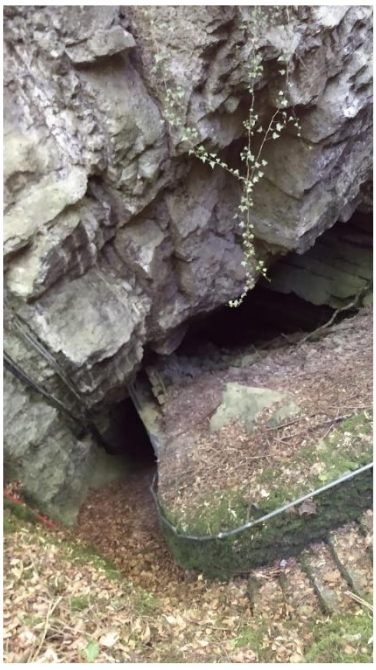

Photo VII.6: The staircase installed in the doline to access the cave (left). The Electrical Resistivity Tomography (ERT) profile of 48 electrodes starts vertically at the bottom next to the staircase (visible on the left) and progresses to the surface where it becomes horizontal (right). Source: ORB (Michel Van Camp).

Photo VII.7: Drip discharge monitoring device within the cave (P1). Source: UCLouvain (Sebastien François).