

**Original citation:**

Castellani, Brian, Rajaram, Rajeev, Gunn, Jane and Griffiths, Frances. (2015) Cases, clusters, densities : modeling the nonlinear dynamics of complex health trajectories. Complexity .

**Permanent WRAP url:**

<http://wrap.warwick.ac.uk/72244>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

"This is the peer reviewed version of the following article: Castellani, B., Rajaram, R., Gunn, J. and Griffiths, F. (2015), Cases, clusters, densities: Modeling the nonlinear dynamics of complex health trajectories. Complexity. doi: 10.1002/cplx.21728, which has been published in final form at <http://dx.doi.org/10.1002/cplx.21728> . This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#)."

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [publications@warwick.ac.uk](mailto:publications@warwick.ac.uk)

warwick**publications**wrap

highlight your research

<http://wrap.warwick.ac.uk>

## Cases, Clusters, Densities: Modeling the Nonlinear Dynamics of Complex Health Trajectories

Brian Castellani · Rajeev Rajaram ·  
Jane Gunn · Frances Griffiths

the date of receipt and acceptance should be inserted later

**Abstract** In the *health informatics* era, modeling longitudinal data remains problematic. The issue is method: health data are highly nonlinear and dynamic, multilevel and multidimensional, comprised of multiple major/minor trends, and causally complex – making curve fitting, modeling and prediction difficult. The current study is fourth in a series exploring a case-based density (CBD) approach for modeling complex trajectories; which has the following advantages: it can (1) convert databases into sets of cases ( $k$  dimensional vectors) based on a set of bio-social variables, called traces; (2) compute the trajectory (velocity vector) for each case, based on (3) key traces from the  $k$  dimensional profile; (4) construct a theoretical map to explain these traces; (5) use vector quantization (i.e., k-means, topographical neural nets) to longitudinally cluster case trajectories into major/minor trends; (6) employ genetic algorithms and ordinary differential equations to create a microscopic (vector field) model (the inverse problem) of these trajectories; (7) look for complex steady-state behaviors (e.g., spiraling sources, etc) in the microscopic model; (8) draw from thermodynamics, synergetics and transport theory to translate the vector field (microscopic model) into the linear movement of macroscopic densities; (9) use the macroscopic model to simulate known and novel case-based scenarios (the forward problem); and (10) construct multiple accounts of the data by linking the theoretical map and  $k$  dimensional profile with the macroscopic, microscopic and cluster models. Given the utility of this approach, our purpose here is to organize our method (as applied to recent research) so it can be employed by others.

---

B. Castellani  
Dept. of Sociology, Kent State University, Ashtabula, OH - 44004  
Tel.: 440-964-4331  
E-mail: bcastel3@kent.edu

R. Rajaram  
Dept. of Mathematics, Kent State University, United States

**Keywords** Longitudinal data · Case-based modeling · Nonlinear dynamics · Complex health trajectories · Differential equations · Vector quantization

## 1 Introduction

Modeling the nonlinear dynamics of complex health trajectories across time presents a number of serious challenges for scientific inquiry [6],[7],[8],[17],[18]. The challenge comes in the form of method (both in terms of the complexities of data and the limitations of conventional techniques).

In terms of data, the challenge is that complex trajectories, be they cohort or longitudinal data: (1) seldom follow a singular common trend; instead, (2) they self-organize into multiple major and minor trends; which, (3) when modeled microscopically, are high dynamic and complex – often taking the form of a variety of complex behaviors – making curve fitting, prediction and control (for example, health management) very difficult; furthermore (4) these continuous trends are often a function of different measurements ( $k$  dimensional vectors) on some profile of biomedical-psycho-social factors and (5) the complex set of qualitative interactions and relationships amongst these variables [6][7][8].

Our last point on data leads to the challenge of technique: (1) while medicine and health are ultimately about the case –  $k$  dimensional vector profiles – health researchers tend to ignore these complex profiles and the set of qualitative interactions of which they are comprised; (2) focusing, instead, on what they deem to be the most salient (and relatively independent) handful of variables relevant to some outcome of concern; (3) which they model by controlling for the remaining profile of variables; (4) furthermore, they study these few factors using some form of linear (and often discrete) modeling/statistics; (5) typically in the search for the most common one or two aggregate trends [6][7][8].

As a result of this approach, there is a major disconnect between health data and health research, making it difficult for scientists to do such things as (1) model the aggregate nonlinear dynamics and complex trajectories of cases or their densities in continuous time; (2) detect the presence of multiple trends (i.e., major and minor) across time; (3) identify and map complex steady-state behaviors (i.e., transient sinks, spiraling sources, periodic orbits); (4) explore and predict the motion of different health trajectories and time instances; or (5) link these different trends to the complex  $k$  dimensional vectors/profiles upon which they are based, so that (6) they can construct a multi-level theoretical model of their topic of study[4,5,17,18].

While the above disconnect between health data and technique is problematic, researchers are beginning to 'turn' to the complexity sciences and computational modeling and related areas of inquiry – genetic algorithms, dynamical systems theory, agent-based modeling, network analysis, differential equations, control theory, etc – for possible solutions [2]. In regards to this

'complexity turn,' we seek to demonstrate the utility of *case-based complexity* for modeling complex health trajectories[3].

Case-based complexity combines case-comparative method with the various theoretical and methodological tools of the computational and complexity sciences to advance the modeling of complex (social and health) systems; which it does by treating complex systems as sets of cases (i.e.,  $k$  dimensional vectors/profiles)[4,5,17,18]. The platform we created for this approach is called the *SACS Toolkit* [4,5,17,18].

The SACS Toolkit is a case based, computationally grounded, mixed methods framework for modeling complex systems. One of its key strengths – called a *case-based density approach* – is its capacity to model the nonlinear dynamics of complex trajectories, particularly in the form of cohort or longitudinal data[17,18]. To do so, it employs a novel combination of case-comparative method in conjunction with vector quantization, genetic algorithms, ordinary differential equations (ODE), Haken's synergetics, the inverse-forward problem, and non-equilibrium statistical mechanics, specifically transport theory and the continuity (advection) partial differential equation (PDE). The result is a *ten-step, multi-level* procedure for transforming the nonlinear dynamics of complex trajectories into cases, clusters and densities[4,5,17,18].

The current study is fourth in a series.[5,17,18] The purpose of the first three papers was to provide a mathematical outline of our approach [5] and to work on several key steps, including (1) a technique for fitting an ODE directly to data and (2) a procedure for using the vector field thus obtained to simulate the evolution of the distribution of cases (as densities) across time using the advection PDE[17,18]. Still, the following remains to be done. We have yet to:

1. assemble our ten steps into a formal outline for others to employ;
2. highlight how the various and multiple outputs of our ten steps go together to create our multi-level model; and
3. demonstrate the utility of our approach in application to several of our recent studies.

Hence, we come to the purpose of the current study: we seek to formalize our case-based density approach, as employed through the SAC Toolkit, in application to several of our recent health studies, including a study on allostatic load [1], public health [6], and international health [18], along with a forthcoming study on depression and wellbeing [19]. While the last three studies are all longitudinal, the first is discrete; nonetheless, we will refer to it here, as it was crucial to developing several of our steps.

## 2 Modeling Health Trajectories: Cases, Clusters and Densities

Case-based density modeling, as employed through the SACS Toolkit, is a *ten-step, multi-level* procedure for studying the nonlinear dynamics of complex trajectories, the process of which can be summarized as follows:

## 2.1 Steps 1 Through 4: Cases, Traces, and Profiles

The purpose of the first four steps is to construct a case-based portrayal of the topic (complex system) of study by: (1) rethinking the database from a case-based (as opposed to a variable-based) perspective; (2) computing the trajectory (velocity vector) for each case, based on (3) an identified set of traces variables (which we will explain below); and (4) constructing a *working theoretical model* to explain the trajectory of these traces.

### 2.1.1 Step 1

The first step is an epistemological one: it requires a cognitive shift from a variable-based to a case-based view of the topic (complex system  $S$ ) of study. According to case-based complexity, cases are complex profiles comprised of a set of inter-dependent variables, which are contextually dependent, nonlinear, dynamic, evolving, self-organizing, emergent, etc ... in short, cases have the same characteristics as a complex system. Theoretically speaking, then, cases (as in a cohort or longitudinal study) can be treated and modeled as complex systems[2,3].

What the case is for any given study, however, can vary significantly. For example, in a public health study we conducted on a Midwestern county in the United States [6], our case was a county, which we conceptualized (using census data) as a set of 20 communities (smaller cases). As such, as shown in Figure 1, for our study we moved back and forth between Summit County (our primary case) and its twenty major communities (our more specific cases).

As a second example, in a recent study we conducted on the negative impact of stress, we treated allostatic load as our primary case [1]. We did this by conceptualizing it as a complex clinical construct, comprised of a large number of interdependent biological subsystems, which are represented by an even larger number of interconnected biomarkers. However, when it came to our database, allostatic load became more of an abstraction, as we did not seek to build a single model of this clinical system. Instead, we sought to treat each of our  $N=1151$  cases as an individualized model. In other words, each case in our study constituted one possible way that allostatic load manifests itself in people's lives; one possible trajectory in the larger state-space of all possible trajectories, based on the unique way allostatic load self-assembles itself for each case.

As these two examples illustrate, it is this unique, case-based approach that distinguishes our method from conventional statistics or conventional mathematical modeling. We begin with the assumption that any complex system of study requires multiple and different (albeit interconnected) models, as there is no one trajectory taken by the system's cases. With this epistemological shift in thinking established, next the database requires further reconceptualization.

From a case-based complexity perspective, each row in a study's database  $D$  becomes a complex case  $c_i$ , where each  $c_i$  is a  $k$  dimensional row vector  $c_i = [x_{i1}, \dots, x_{ik}]$  and where each  $x_{ij}$  represents a measurement on the profile of

longitudinal variables for  $D$  – what case-based researchers call the *case profile*. For example, in our public health study – see Figure 2 – we treated each of the 20 cases (communities) as a set of measurements on an in-depth profile ( $k$  dimensional row vector) of contextual, compositional and health factors. Furthermore, these variables were measured at two discrete time points (2000 and 2010). In turn, in a recent international health study, we examined the longitudinal relationship between per-capita GDP and human longevity rates (our two profile variables) for 156 countries (each a complex dynamical case) over the course of 63 years. Data for this model came from the widely used Gapminder dataset (<http://www.gapminder.org/>).

The temporal nature of these two examples take us to our next point: cases  $c_i$  are not static; instead, they are dynamic and evolving. As such, in terms of cohort and longitudinal databases  $D$ , each case  $c_i$  in  $D$  is, ultimately, a *complex dynamical system*  $c_i(j)$ , where  $j$  denotes the time instant  $t_j$ . In turn, if the trajectories of cases  $c_i$  change across time/space, so too must their vector configurations  $c_i = [x_{i1}, \dots, x_{ik}]$ . As such, in terms of cohort and longitudinal studies,  $D$  is comprised of a series of  $c_i(j)$ , one for each moment in time/space  $t_j$  (discrete or continuous), on which a set of measurements are taken to construct a particular model of the complex system of study  $S$ .

Two examples: First, in a new study we are conducting on depression and wellbeing, we examined a total of 84 monthly time-stamps, for a total of seven years worth of trajectory data [19]. Data for this study (N=259 cases) came from a subsample of the *Diamond Prospective Longitudinal Cohort Study*, one of the largest primary care depression cohort studies worldwide. Second, in our international health study (as mentioned above) we examined the longitudinal relationship between per-capita GDP and human longevity rates for 156 countries over the course of 63 years. In fact, Figure 3 shows a microscopic model of the trajectories of these countries across time (shown in blue) along with the model we ‘fitted’ to the data (shown in green). The X-axis represents GDP; and the Y-axis represents life expectancy.

### 2.1.2 Steps 2 and 3

With the database reconfigured into a case-based framework, the next two steps are to identify and model the key traces of the system.

The challenge with modeling cohort and longitudinal data is that, given some complex profile of study, the resulting vector configuration  $c_i = [x_{i1}, \dots, x_{ik}]$  and corresponding vector field are  $k$  dimensional and therefore too complex or dynamic to be accurately modeled [2, 16]. As a result, even with a working map in hand, one rarely has direct access to the actual state of the system studied. Instead, *pace* observability and control theory, one studies the system’s state in a modified form.

Following Byrne and Callaghan [2], we refer to this modified form of the system as its *trace*. In other words, while the ultimate modeling goal of case-base complexity is idiographic analysis, one never fully models the complete

complexity of a case or set of cases; instead, one only studies their traces, albeit from a case-based, complex systems perspective!

Theoretically speaking, the trace can be anything measurable that is directly/indirectly influenced by the actual state of a system. In practice, however, we find it useful (at least for our case-based density approach) to begin with the output/dependent variables in a case-based profile. Two reasons:

First, the output is typically what researchers are trying to understand, model, manage or control. For example, in our study of public health, we were primarily interested in community-level health outcomes across time (See Figure 2); and, in our study of international health, the focus was on human longevity (across time) in each of our 156 countries (See Figure 3).

Nonetheless, once these output traces are explored, one continues onward to increase the complexity of the study by exploring their intersection with other key traces – which takes us to the next point.

Second, it provides a useful way to manage the complex causality residing within the case-based profiles. As we will discuss later, the purpose of Step 10 is to explore how the traces (as outcomes, outputs, dependent variables) link to, evolve or change in relationship to other key biological, psychological, social or ecological traces. For example, in our public health study, we examined our community-level health outcomes in relation to the compositional and contextual factors shown in Figure 2; and, in our international health study, we examined human longevity rates in each country in relation to that country's per-capita GDP.

With the traces identified, the next thing is to use these them to compute the *velocity vector* for each case. Later, in Steps 5 and 6, we detail the process of computing the traces. Here we want to provide our rationale for why velocity vectors are so important to our approach.

A key feature of our approach, mathematically speaking, is our link between an 'algebraic-based' definition of cases as  $k$  dimensional vectors and a 'calculus-based' definition of vectors as quantities with direction and magnitude. We make this link for several reasons: it allows us to (1) treat cases as continuous trajectories (traces); (2) compute these continuous trajectories (traces) as a function of change between time-stampss (something statistics struggles, at best, to do); (3) examine these changes as a function of discrete or continuous measures on the  $k$  dimensional vector for each case (which brings in our theoretical model); (4) employ ODEs to explore the rate of change or velocity of cases (first-order ODE), as well as acceleration (second-order ODE) if needed; and (5) link steps 1 through 4 with the modeling processes in steps 5 through 10. As such, computing the velocity vector for each case functions as *the* main methodological link upon which our approach is based.

### 2.1.3 Step 4

With the velocity vectors computed, the next step is to construct a working theory (map) of the topic of study. The purpose of this map is to theorize, albeit tentatively, how the factors in the case-based profile – as a complex

system of interacting variables – go together in relation to the set of outcome(s) being observed; which are treated as trajectories (traces) across time/space.

For example, to arrive at our conceptualization of community health as a complex system we needed a theoretical map. The result was Figure 4. The utility of this map is that it gave us an idea of what traces to explore and how to think about their complex inter-relationships, particularly in relation to our community health outcomes (which we will discuss later). For example, looking at the map, one sees the key factors outlined in Figure 2 (compositional and contextual) as well as some of the key environmental forces we explored in our study; also one sees the three types of maps we constructed for our study, including (as we will discuss in Step 10) a social network analysis of the relationships amongst our 20 communities.

Another example of the importance of a theoretical model is study on allostatic load. As shown in Figure 5, the utility of this map was that it gave us an idea of what traces to explore and how to think about their complex inter-relationships, particularly in relation to various health risk outcomes (which we will discuss later). Using this map, we worked with context experts to settle on 20 key biomarkers – each constituting one of the key variables in our  $k$  dimensional profile. We then factor analyzed these 20 biomarkers to construct a seven-factor solution, as shown in Figure 6. In turn, these seven factors became our seven main trace variables.

## 2.2 Step 5: Major and Minor Clusters and Trends

With the theoretical map constructed, the fifth step is to identify the major and minor trends in the data, based on the traces initially chosen for study. This step is done using k-means cluster analysis and the topographical neural net known as the Self-Organizing Map [10][11][13][12]. Our usage of k-means and the SOM is unique in three important ways:

### 2.2.1 Knowledge Free Clustering

First, we take a *knowledge-free* approach to trend identification. Based on the multiplicity of possible trajectories generated by the theoretical model from Step 4, we do not make any reductive assumptions about the number of cluster trajectories or possible major and minor trends in the data. Instead, we strive to identify these trends first and then model them separately for each trace, thereby allowing for the creation of multiple models for the same system.

For example, in our study of depression and physical wellbeing, we identified 18 different cluster trajectories. And, in our public health study, we arrived at a seven-cluster solution. Finally, in our allostatic load study, we identified nine clusters, which are shown in Figure 7.

Despite the differences in outcome, in both studies the goal was the same: to allow the data, through the key traces identified, to interact with, speak to, temper, impact, disagree with, modify, or corroborate the theoretical model.



The model that proved the best fit, based on its corroboration with our theoretical map, is the one used.

That is not, however, where the search for trends needs to stop. Depending on how complex the case-based profile under study is, we can go on to identify sub-trends (and even sub-sub-trends) for any particular cluster, thereby giving us a hierarchy of models that range from a single model for the entire database of cases (which will hardly capture any complexity at all) all the way down to a model for each individual case (which captures, perhaps, too much complexity).

It is because of our *knowledge-free* approach to clustering that our method is a data-driven special case of the inverse-forward problem: we start with data analysis to identify trends; then we move to and develop the theoretical model to organize the causal mechanisms for the trends; then we go back to the data to identify sub-trends if need be and also model the mechanisms identified; and then back to the theoretical model. As the data grows in size with the addition of cases or time-stamps, we can repeat the process, hence the method scales as well.

### 2.2.2 Longitudinal Clustering

Second, we use vector quantization to engage in longitudinal clustering. We need to emphasize that Step 5 involves clustering case trajectories; not static profiles, as is done in traditional clustering. To cluster cases longitudinally, we treat each time instance as a measure, and the total of time instances/measures as the longitudinal  $k$  dimensional vector profile for each case. In turn, these trajectories can be combined (appended to one another) so that the cluster solution is based on similarities in evolution across all of the trace trajectories. For example, in our study of international health we appended the trajectory for per-capita GDP with the trajectory for longevity rates for each of the 156 countries in our database.

### 2.2.3 Clustering to Corroborate

Third, we use k-means and the SOM as a method of corroboration. K-means is a partitional (as opposed to hierarchical) iterative clustering technique that seeks a single, simultaneous clustering solution for some proximity matrix. For k-means, reference vectors are centroids, representing the average for all the cases in a cluster. The SOM is a topographical artificial neural network that maps high-dimensional data onto a smaller, three-dimensional space, while preserving, as much as possible, the complex patterns of relationships amongst these data. For the SOM, reference vectors are actual points, neurons, which represent the weighted average of the cases clustering around it. Both k-means and the SOM are forms of unsupervised learning, as cluster membership is not known ahead of time.

In terms of a case-based density approach, these methods are used in combination as follows: k-means is used first because it requires that the number

of centroids be identified ahead of time, based largely on some rationale, even if tentative or conjectural. As shown in Figure 7, following convention (and as discussed earlier), the goal of multiple runs is to find a solution that fits the data well and resonates with our theoretical model, even when exploratory.

Next, the SOM is run. Because the SOM is entirely unsupervised, if it arrives at a solution similar to the k-means this provides an effective method of corroboration. The closer the final quantization error and final topographic error are to zero, the better the fit of the model.

The SOM graphs its cluster solution onto a variety of three-dimensional, topographical maps. The three we typically use are the u-matrix, eigenvector map, and components map. On the u-matrix and eigenvector maps, cases most like one another are graphically positioned as nearby neighbors, with the most unlike cases placed furthest apart. Both maps are also topographical: valleys, or darker colored, areas are more similar in profile; while hilly, or brighter colored areas, are more distinct. The component maps (which we will discuss in Step 10) visualize how each of the variables (traces) from the complex profile of study contribute to the final cluster solution and to the positioning of cases on the u-matrix and eigenvector map.

A good example of this output comes from our study on depression and physical wellbeing. As Shown in Figure 8, Map A and Map B are graphic representations of the cluster solution arrived at by the Self-Organizing Map (SOM) Neural Net, referred to as the U-Matrix. In Figure 8, Map A is the three-dimensional (topographical) u-matrix: for it, the SOM adds hexagons to allow for visual inspection of the degree of similarity amongst neighboring map units; the dark blue areas indicate neighborhoods of cases that are highly similar; in turn, bright yellow and red areas, as in the upper right corner of the map, indicate cases that are very different from the rest. Map B is a two-dimensional version of Map A that allows for visual inspection of how the SOM clustered the individual cases. Cases on this version of the u-matrix (as well as Map A) were labelled according to their k-means cluster membership (the 18 cluster solution we discussed earlier) to see if the SOM arrived at a similar solution, which (roughly speaking) it did.

### 2.3 Microscopic and Macroscopic Models

As mentioned in the abstract, to construct our microscopic model (Steps 6 and 7), we employ a combination of genetic algorithms and ODEs; and to construct our macroscopic model (Steps 8 and 9), we employ the continuity (advection) PDE in application to the vector field generated by our microscopic model. As such, before moving on to our next set of steps, a bit of detail on the mathematics behind them is necessary. Let us start by considering the following ordinary differential equation (ODE):

$$\dot{x} = f(x), \quad x(0) = x_0, \quad (1)$$

where  $x \in X \subset \mathbb{R}^K$  is a compact set. We denote the solution of this ODE (1) by  $\phi_t(x)$ , starting from the initial condition  $x$ . We use this ODE (1) to define two linear infinitesimal operators,  $\mathcal{A}_K : L^2(X) \rightarrow L^2(X)$  and  $\mathcal{A}_{PF} : L^2(X) \rightarrow L^2(X)$ , which are defined as follows:

$$\mathcal{A}_K \rho = f \cdot \nabla \rho,$$

$$\mathcal{A}_{PF} \rho = -\nabla \cdot (f \rho).$$

The domains of the above operators are given as follows:

$$D(\mathcal{A}_K \rho) = \{\rho \in H^1(X) : \rho|_{\Gamma_o} = 0\},$$

$$D(\mathcal{A}_{PF} \rho) = \{\rho \in H^1(X) : \rho|_{\Gamma_i} = 0\},$$

where  $\Gamma_o$  and  $\Gamma_i$  are the outflow and inflow portions of the boundary  $\partial X$  defined as follows:

$$\Gamma_o = \{x \in \partial X : f \cdot \eta > 0\},$$

$$\Gamma_i = \{x \in \partial X : f \cdot \eta < 0\},$$

where  $\eta$  is the outward normal to the boundary  $\partial X$ . The semigroups corresponding to the  $\mathcal{A}_K$  and  $\mathcal{A}_{PF}$  are called as Koopman ( $\mathbb{U}_t$ ) and Perron-Frobenius ( $\mathbb{P}_t$ ) operators respectively. These operators are defined as follows:

$$\mathbb{U}_t : L^2(X) \rightarrow L^2(X), \quad (\mathbb{U}_t \rho)(x) = \rho(\phi_t(x))$$

$$\mathbb{P}_t : L^2(X) \rightarrow L^2(X), \quad (\mathbb{P}_t \rho)(x) = \rho(\phi_{-t}(x)) \left| \frac{\partial \phi_t(x)}{\partial x} \right|^{-1}$$

where  $|\cdot|$  denotes the determinant. These semigroups can be shown to satisfy the following partial differential equations [14]:

$$\frac{\partial \rho}{\partial t} - \mathcal{A}_K \rho = 0, \rho|_{\Gamma_o} = 0; \quad \frac{\partial \rho}{\partial t} - \mathcal{A}_{PF} \rho = 0, \rho|_{\Gamma_i} = 0.$$

Of the two, we use the Perron-Frobenius semigroup and the corresponding advection PDE. Here, again, at a greater level of detail, is why: although the solution trajectory  $\phi_t(x)$  for each finite dimensional case  $x \in \mathbb{R}^K$  (given  $\mathbf{f}(x)$ ) is, in general, nonlinear, the motion of the density of cases (governed by the Perron-Frobenius semigroup  $\mathbb{P}_t$ ) is (a) linear and (b) acts on the infinite dimensional space  $L^2(X)$ . In other words, the advection PDE (2) is solved by  $\mathbb{P}_t$ . In turn,  $\mathbb{P}_t$  functions as a conduit for simulating and watching the macroscopic evolution of ensembles of trajectories, given the vector field  $\mathbf{f}(x)$ , which governs the nonlinear microscopic motion of each individual case.

Here, then, are the microscopic and macroscopic models upon which our approach is based:

$$\begin{array}{c}
\boxed{\text{Microscopic model for nonlinear evolution of each case trajectory}} \\
\Downarrow \\
x' = \mathbf{f}(x); x(0) = x_0; x \in X \subset \mathbb{R}^K \\
\boxed{\text{Macroscopic model for the linear evolution of densities of cases}} \\
\Downarrow \\
\rho_t + \nabla \cdot (\rho \mathbf{f}) = 0; \rho|_{\Gamma_i} = 0; \rho(x, 0) = \rho_0(x). \tag{2}
\end{array}$$

#### 2.4 Steps 6 and 7: The Microscopic Model and Steady-State Behaviors

In addition to the mathematics upon which they are based, Step 6 and Step 7 involve a rather complicated set of procedures, which we have outlined in detail elsewhere. Our goal here is to provide a quick overview of the procedures involved in completing them. For more details see [17, 18].

##### 2.4.1 Identify the Data-Driven Vector Field

In terms of constructing the microscopic model, the form of the vector field  $\mathbf{f}$ , which is a part of the ODE (1), is completely unknown. In other words, we do not have a pre-conceived function for the vector field model. As such – and for a second time – we employ our *knowledge-free* approach to modeling: this time looking for the best fit among polynomials of arbitrary degree using genetic algorithms. Our attempt to fit a curve to this data-driven vector field constitutes another of the novel aspects of our approach.

To obtain a vector-field model for the velocities of the traces, we use a curve fitting algorithm to fit each case trajectory with a smooth curve; which we then differentiate with respect to time. Our purpose in doing so is to find the instantaneous (continuous) velocities for the time instants provided by the data – which we discussed above in Step 2. A discrete velocity vector  $\mathbf{f}(x_k)$  is thus obtained, situated at each of the cases  $x_k$ .

##### 2.4.2 Obtaining the Microscopic Model

With this discrete vector field obtained, we then fit a smooth continuous vector field  $\mathbf{f}(x)$  of velocities. To do so, we use a genetic algorithm. The utility of a genetic algorithm is its 'brute force' and evolving power to fit complex dynamical databases with a useful ODE model, while (at the same time) preserving its complexity. We choose to fit polynomial vector fields as they are well known to capture much of complex dynamics and they are dense in most function spaces that are of practical interest such as  $L^2(X)$ .

### 2.4.3 Validity Check

With the above complex, next the ODE (1) is then solved for the initial conditions  $x_0$  from the database. Each trajectory from the ODE model, thus obtained, is compared with the trajectories of the data for error. This process is repeated, if necessary, after re-clustering any minority trends in a separate cluster, from the aggregate trajectories. After iterations, a good model of the major and minor trends of case trajectories in time is thus obtained.

### 2.4.4 Searching for Complex Steady-State Behaviors

The utility of the microscopic model is that, unlike the cluster model, it is devoid of cases; constituting, instead, the data-driven space of all possible trajectories. Equally important, it can be visualized as a movie across all instantaneous time-stamps in the database. As such, the model can be visually inspected to identify important steady-state behaviors and to note the manner in which the trajectories evolve across continuous time, including changes in velocity.

For example, Figure 9 shows the state-space for our depression and physical wellbeing study, which included 84 monthly time-stamps across a 7-year period of time. Ten time-stamps are shown, beginning with four time-stamps from the first year (3 months, 6 months, 9 months and 12 months) and then one time-stamp for each subsequent year – each constituting the point at which new data were collected. Looking at the time-stamps, one sees (1) the emergence of stable spiraling equilibrium points marked in red, which enter the relevant regions of the state space for time-stamps 3 through 24; (2) an unstable spiraling equilibrium point appearing and leaving the state space between time-stamp 36 and 84; and (3) a saddle appearing at time-stamp 84.

Because, in our approach, the ODE that models the evolution of depression trajectories is non-autonomous  $\dot{x} = \mathbf{f}(x, t)$ , time  $t$  is an independent variable and hence the vector field changes its nature as time evolves, as seen in the emergence and disappearance of various steady-state and transient behavior (such as rifts etc.). This is one of the main advantages of using an ODE to model the trajectories: both the steady-state and transient behaviors of trajectories can be studied with time as an independent variable (which allows for change of these behaviors across time) by using the multiple ODE models (both autonomous and non-autonomous) that the genetic algorithm fits as possible explanations of the trajectories from the standpoint of ODEs.

## 2.5 Steps 8 and 9: The Macroscopic Density Model

With the cluster and microscopic models complete, the next stage is to assemble the macroscopic model, which involves two key steps.

### 2.5.1 Simulating the Transport of Densities

First, we take the vector field  $\mathbf{f}$  from our microscopic model – which is governed by the ODE – and use the advection PDE to translate it into the macroscopic motion of case-based densities. In doing so, we add a third level to our approach, focused on the macroscopic, non-equilibrium properties of the system as a whole.

Our approach is motivated by thermodynamics [15], where the state of the system is the characteristic of a density of particles and their properties (as in the case of temperature or pressure) rather than the individual particles themselves. As we discussed earlier, the major challenge of modeling longitudinal data is that trajectories are often highly complex. For example, in our study of international health, our microscopic model – while useful for identifying major and minor trends and steady-state behaviors – was still incredibly dynamic (See Figures 3 and 9). In such instances, our approach is useful because, following Haken [9] and others [28,30], it models the ensemble of cases as the macroscopic movement of densities across continuous time; which are, generally speaking, lower dynamic and therefore easier to model for common patterns and trends across time.

The key aspect of the advection PDE is that it models the transport of a physical quantity according to a given vector field  $\mathbf{f}$  (as in the case of our microscopic model) in addition to conserving the physical quantity itself. The dynamical state of the advection PDE is a density  $\rho$ , which is a function of both space  $x$  and time  $t$ . In turn, the density function  $\rho$  is basically the physical quantity per unit area in two dimensions.

Given an initial distribution of cases  $\rho_0(x)$ , the advection PDE simulates the evolution of  $\rho_0$  under the assumption that (a) the total number of cases remains the same, and (b) each case  $x$  moves according to the velocity vector  $\mathbf{f}(x)$ . The boundary condition  $\rho|_{\Gamma_i} = 0$  ensures that no new cases enter the state space through the inflow portion of the boundary given by

$$\Gamma_i = \{x \in \partial X : \mathbf{f}(x) \cdot \boldsymbol{\eta} < 0\}, \quad (3)$$

where  $\boldsymbol{\eta}$  stands for the outward normal on that boundary  $\partial X$ . And, we use the vector field  $\mathbf{f}$  obtained in microscopic model to simulate the advection equation given in (2) along with the boundary condition (3). As a final note, the validity check for the motion of densities is mathematically inherent because the vector field  $\mathbf{f}$  is already checked in the microscopic model. (For more details on our approach, see [17,18].)

### 2.5.2 Simulating Initial and Novel Conditions

With the macroscopic model built, the next step (Step 9) is to run it – which we do by introducing known or novel sets of initial conditions. Unlike the microscopic model, which is devoid of cases, the macroscopic movie re-introduces different distributions of cases back into the model to explore their *actual* trajectory amongst all *possible* trajectories in the microscopic model.

For us, these initial conditions come in three types: (1) conditions based on the initial dataset upon which the microscopic vector field  $\mathbf{f}$  was based; (2) conditions based on the major and minor trends identified in the cluster and microscopic models; and (3) novel conditions researchers wish to explore, based on the results from running conditions (1) and (2).

Simulating these 'various' initial conditions is important to our approach because it brings us full circle, moving us from the inverse to the forward problem in physics: in other words, while we use the microscopic model  $\mathbf{f}$  to simulate the macroscopic evolution of densities (using the advection equation), we do so by returning to the initial conditions of the raw data (be it known or novel) for corroboration.

For example, Figure 10 provides a series of snapshots from a simulation we made for our international health study. In terms of reading Figure 10, the x-axis represents GDP and the y-axis represents life expectancy; also, scores on the axes were converted to z-scores for normalization and comparison. In this example, the initial conditions  $t = 0$  (which are shown in Models A and B) were based on the original Gapminder dataset (<http://www.gapminder.org/>). (The complete movie can be found at [http://www.personal.kent.edu/~bcaste13/macroscopic\\_model.mp4](http://www.personal.kent.edu/~bcaste13/macroscopic_model.mp4).)

In contrast, Figure 11 provides a series of snapshots from a simulation we made for our depression and physical wellbeing study. In this simulation, five different sets of initial conditions we explored, based on key trends identified in the cluster and microscopic models. As with Figure 10, scores were converted to z-scores, with the y-axis representing physical wellbeing; and the x-axis representing depression.

As these two examples illustrate, our macroscopic model adds several advantages to our multi-level, case-based approach to modeling complex health trajectories.

1. To begin, different regions of the simulation can be explored to see how different sets of cases evolve (speed up, slow down, spread out, condense inward toward the center of the density, etc) across time.
2. And, these movements can be calibrated using a number of indicators, such as the the contour plot of speed (magnitude of velocity) – as shown in the lower right graph in Figure 10.
3. Also, the non-equilibrium clustering of trajectories during transient times can be studied by looking at the Lyapunov density plot. In Figure 11, for example, high values in the Lyapunov density plot (shown in the upper right graph) indicate that a large number of trajectories have squeezed through that region in the state space.
4. We can also use these simulations to predict the longitudinal evolution of cases across time and space.
5. And, we can study the complexity of various transient case dynamics, which we do by stopping the evolution of the model prior to some key moment in the simulation.

6. Also, based on the exploration of various novel conditions, predictions can be made for the evolution of case profiles and time instants that are not part of the database.
7. And, multiple models can be tested simultaneously to find the model that best explains the data.
8. Finally, new data can be incorporated with ease into the modeling process, thereby providing us with a means to improve the model's fit and predictive value in response to the database's evolution, expansion, development, etc.

## 2.6 Step 10: Constructing Multiple Accounts

Consistent with the overarching theme of case-based complexity – which seeks to find differences through idiographic, case-comparative analysis – our approach, once again, distinguishes itself from convention. In Step 10, we do not seek to build a single causal model or overarching account of the data. Instead, we seek to construct *multiple models, multiple accounts*. And, we aim these multiple accounts at explaining (exploring, understanding, etc.,) key differences, (distinctions, variations, nuances etc.,) in the data.

Equally important, we assume that these multiple accounts come from the theoretical map – as constructed in Step 4 – and its  $k$  dimensional profile of traces. And we assume – as discussed in Step 1 – that this map and profiles are best studied as complex systems. As such, while it is useful to explore variable-based trends across data; in our approach the emphasis is on the intersection of traces and their self-organizing and emergent impact on some initial (dependent / outcome) trace of concern.

With our map and case-based profile in hand, we construct our multiple accounts by employing the following two-stage process:

### 2.6.1 Corroboration of the Three Models

During the course of completing steps one through nine, a significant amount of information is generated. It is therefore necessary, as a first course of action, to summarize and further corroborate these data into a *multi-level, working narrative* to identify key issues for which we seek to develop an account.

**1. The Cluster Model:** An easy place to start is with the cluster model. Here, the goal is to verify further the veracity of the cluster trajectories and the major and minor trends they represent. Such verification includes working with context experts to:

1. determine if the clusters make clinical or theoretical sense.
2. examine if certain clusters need to be discarded or combined to create a larger cluster.
3. determine – as discussed in *Step 2.2.1* – if further sub-clustering is necessary or clinical or theoretically meaningful.
4. name the final cluster and sub-cluster solutions, as well as major and minor trends, based on how their trajectories differ from one another.



5. assemble all this information into a working narrative.
6. And, finally, identify key issues for which an account of this narrative is required, including hypothesizing how the other  $k$  dimensional traces in the theoretical model and case-based profile might account for these differences.

For example, in our study on depression and physical wellbeing, the two lead clinicians (both authors on the current paper) had to pour over the data to make sure everything made clinical sense, including tentatively hypothesizing how the other traces in the theoretical model might account for these differences.

The same was true of our allostatic load study, albeit at an even greater level of detail, as the clusters for this study were based on our factor analysis (we discussed this earlier); which was, in turn, based on our 20 key biomarkers. As such, the biologists and clinicians on our team had to corroborate the biomarker linkages found in our cluster solutions – shown in Figure 7 – to make sure they fit with what they and the literature knew to be biologically true or possible.

In turn, the nine clusters shown in Figure 12 also had to make sense in terms of our theoretical model of allostatic load (See, from earlier, Figure 5). For example, Map C in Figure 12 is a graphic representation of the relative influence that the seven traces (shown in Figure 7) had on the SOM cluster solution. The SOM generated a mini-map for the seven traces, each of which can be overlaid across maps A and B. Each of these mini-maps was also visually inspected to examine what its rates were across the different neighborhoods (clusters of cases – See Map B, Figure 12). Dark blue areas in Map C indicated the lowest rates for a factor; and the bright red areas indicated the highest rates for a factor. For example, looking at the map for Factor 6 (Blood Sugar), its rates were extremely low across most of the map, except for the lower right corner, where (looking at Map B) the SOM placed Cluster 6.

**2. The Microscopic Model:** Next, the insights from the cluster model need to be integrated and further corroborated with the microscopic model. Such integration and corroboration includes working with context experts to:

1. verify the clinical and theoretical utility of the various transient steady-state behaviors initially identified during *Step 2.4.4* of the model building process.
2. use the cluster trajectories to visually corroborate the manner in which the microscopic vector field evolves across continuous time.
3. examine how key velocity change moments in the microscopic model coincide with key changes in the trajectories of various major and minor clusters. This is done by (a) identifying the region in the state-space where the behavior occurs; (b) studying the clusters corresponding to the regions; and (c) watching the movies and the actual trajectories from data (both overall and at the interesting time instants) to see if they corroborate one another.

4. use the findings from the cluster and microscopic model to identify different regions of the state-space that are qualitatively important, and to then label them accordingly.
5. and, finally, hypothesize how the other  $k$  dimensional traces in the theoretical model and case-based profile might account for these differences in the state-space and their respective evolution across time-space.

For example, as shown in Figure 9, in terms of our study of depression and physical wellbeing, we identified five major regions into which we could fit all 18 of our clusters, based on the major and minor trends we noted in our cluster solution. Roughly speaking, these regions approximated how physical wellbeing and depression work together as a differential diagnosis for sorting clusters according to their key differences. Figure 11 – which we already mentioned but will discuss further in a moment – provides the name of these five major regions. In addition, we confirmed that the transient steady-state behaviors from our microscopic model (Figure 9) coincided with key shifts in the cluster trajectories from our cluster model.

**2. The Macroscopic Model:** The last part of our corroboration process is to add the macroscopic model to our multi-level narrative by integrating the insights from its density simulations. Such integration and corroboration includes working with context experts to:

1. confirm (both clinically and theoretically) the utility of the results gained in *Step 2.5.2* of the model building process, which includes validating the various novel scenarios simulated, based on the results from the microscopic and cluster models.
2. verify *how* and also *when* the transient steady-state behaviors identified in the microscopic model manifest themselves in the density model, given that both models are based on the same vector field generated by the ODE.
3. and, finally, hypothesize how the other  $k$  dimensional traces in the theoretical model and case-based profile might account for these differences in the density model.

For example, looking at Figure 11 from our depression and physical wellbeing study, we see two snapshots –  $t = 0$  months and  $t = 84$  months – which we labeled according to the results from our cluster and microscopic models. During the course of exploring these five simulations, one of the findings (amongst many) that stood out was how the density plot for the healthiest region (lower right) changed across time, moving more toward the center. We also noted how this healthy plot shifted and became more distributed across time, with the upper right side stretched more toward increased depression, and the lower right stretched more toward decreased physical wellbeing. The question we sought to answer, based on this finding, and for which we engaged in a series of hypotheses, was why? Answering this question takes us to the final stage of our model building process: constructing multiple accounts.

### 2.6.2 Constructing Multiple Accounts

The final stage of the modeling process is to construct a series of accounts (causal models) that help make sense of the complex health trajectories studied. Again, no one account is expected to explain everything. Instead, we seek multiple accounts, multiple explanations.

To do so, we begin with method. Unlike the previous steps, however, this last stage need not follow any particular methodological protocol. In fact, in our work we have employed a variety of methods, including statistical, qualitative and historical analysis. Of these methods, however, perhaps the most useful is to re-run steps five through nine.

As a quick overview: starting with the cluster model, one would proceed, as required, to examine a new set of traces. However, this time there is an added dimension, as the goal is to construct an account of how the nonlinear dynamics of these additional traces coincide, interconnect with, influence or impact the original outcome traces. For example, in our depression and health study, we were specifically interested in how the complex intersection of employment, income and negative life events impact – across time – our 18 cluster trajectories, slowly stitching these new traces together to form a complex model of these data.

Nonetheless, often one does not have the luxury of continuous data; or researchers may be concerned with traces of a different type, as in the case of discrete, qualitative or historical data; or they may be interested in other methods, as in the case of network and agent-based modeling.

For example, in our public health study, we used multiple linear regression and the unstandardized F-scores from our k-means cluster analysis to determine the relative impact that various compositional and contextual traces had on our community-level health outcomes.

In this same study we also used a combination of qualitative and historical data. For example, to examine the views and opinions of people living in the poorer communities surrounding the two urban centers in our study, we turned to a series of focus groups that local public health researchers had done. And, to make historical sense of how out-migration impacted community-level health, we turned to a local newspaper series on access to healthcare.

Going even further, we employed the tools of complex network analysis, examining (*pace* Christakis and Fowler’s work on obesity networks) how changes in communities in one part of our county-wide network influenced changes in another. Figure 13, for example, is a network representation of our cluster analysis data. The network is made up of the seven clusters we identified in our study, labeled 1 through 7. Around each cluster are the communities associated with it.

In terms of reading this network, the greater the distance between cluster centers, the less alike these clusters are in health and economic wellbeing; and, the greater the distance a community is from its cluster center, the less similar its configuration is to the other communities in its cluster. One of the questions we examined using this network was: given changes in the overall economic

wellbeing of all twenty communities, which of the poorer communities fell further into a poverty trap, relative to the rest? The answer was found in the bottom two clusters, both 2 and 7, with the communities in Cluster 7 having the poorest, overall, health outcomes.

And, as further evidence of this 'poverty-trap' effect, we constructed an agent-based model. Our goal in doing so was to test our data-driven assumption about the role out-migration played in the different community-level health outcomes we found. Figure 14, for example, shows (*pace* Schelling's segregation model) that, as more affluent communities pull away from the poorer communities, the latter became segregated into poverty traps, identified by the lighter grey clusterings in the center picture of our Netlogo model.

Still, despite these differences in method and technique, when it comes to the final step of our case-based density approach, two things are consistent in the *account building* process. The first has to do with focus: no matter what the technique used, the purpose is to construct multiple accounts that help explain the nonlinear dynamics of complex longitudinal health data. More specifically, this means making sense of:

1. the different cluster trajectories;
2. major and minor trends;
3. various transient, steady-state behaviors;
4. the movement (across space/time) of various density distributions;
5. and, the evolution and dynamics of different regions of the state-space, including:
  - (a) differences in the speed and velocity of key trajectories; and
  - (b) the prediction of different known or novel density outcomes.

The second has to do with the preliminary division of the data. Regardless of the method used, before one constructs any account, the data need to be divided according to each case's cluster membership and the corresponding trend or region of the state-space to which it belongs. In other words, it is necessary to divide the database into separate sub-databases according to the different clusters or trends to which cases belong. Once divided, one can then explore these clusters and trends separately and in comparison to one another, looking for different patterns within and also across clusters and trends.

With the construction of these different and multiple accounts complete, one has reached the end of the modeling process – that is, unless, as a function of the study, new or novel additional traces or cases are included, and therefore further modeling is required.

### 3 Conclusion

In this paper, we have brought together the various steps outlined in our previous research[17,18] to organize them into a single method, called a *case-based density approach* (CBD) – which involves ten steps.

In terms of summarizing these ten steps, our main thesis is that complex longitudinal data are inherently multi-level, case-based systems that manifest themselves from the *bottom-up* (as the microscopic, high-dynamic behavior of individual cases) as well as the *top-down* (as the macroscopic, low-dynamic behavior of densities).

Our secondary thesis is that to model such complex case-based, longitudinal data, researchers need to acknowledge that singular one-size-fits-all models are not sufficient; instead, new and multiple models are necessary. Furthermore, these multiple accounts need to be data-driven and predictive (if only in the short range).

Third, we acknowledge that complex phenomena cannot be perfectly (or often even directly) modeled using mathematical models. Instead, one typically studies the traces of a system's complexity. Related, these traces are often unknown; and only identified through multiple exchanges with subject matter experts. However, modeling the traces of complexity is the first step towards understanding a system's causal mechanisms, which is what we have endeavored to achieve with our approach.

Given these three main points, to model complex health trajectories it is necessary to draw upon a wide variety of concepts and techniques from across the complexity sciences, including the ideas of non-equilibrium statistical mechanics, transport theory and thermodynamics. To model the motion of density of cases, we specifically employed the advection PDE, which serves as a conduit to translate the microscopic motions modeled by the vector field  $\mathbf{f}$  into the macroscopic evolution of the density  $\rho$ , while preserving the number of cases intact. This is a very novel feature of our approach, which has yet to be used to observe the complex behavior of health trajectories longitudinally in time.

Finally, Step 10 is a very important part of the modeling process, as it seeks to provide a series of accounts of the causal mechanisms that drive the nonlinear dynamics of health trajectories. As such – and as mentioned earlier – we are currently studying a large sample of the *Diamond Prospective Longitudinal Cohort Study* to develop further this step. This is the future work that is currently underway.

## References

1. Buckwalter, J., Castellani, B., McEwen, B., Karlamangla, A.S., Rizzo, A.A., John, B., O'Donnell, K., Seeman, T.: Allostatic load as a complex clinical construct: A case-based computational modeling approach. forthcoming in *American Journal of Human Biology* (2015)
2. Byrne, D., Callaghan, G.: *Complexity Theory and the Social Sciences: The State of the Art*. Routledge (UK) (2013)
3. Byrne, D., (eds.), C.C.R.: *The Sage handbook of case-based methods*. Sage Publications Ltd. (2013)
4. Castellani, B., Hafferty, F.: *Sociology and Complexity Science: A New Field of Inquiry*. Springer, Germany (2009)

5. Castellani, B., Rajaram, R.: Case-based modeling and the sacs toolkit: A mathematical outline. *Computational and Mathematical Organizational Theory* **18(2)**, 153–174 (2012)
6. Castellani, B., Rajaram, R., Buckwalter, J.G., Ball, M., Hafferty, F.W.: *Place and Health as Complex Systems*. Springer Briefs on Public Health, Germany (2014)
7. Feng, Q.Y., Griffiths, F., Parsons, N., Gunn, J.: An exploratory statistical approach to depression pattern identification. *Physica A: Statistical Mechanics and its Applications* **392(4)**, 889–901 (2013)
8. Gunn, J., Elliott, P., Densley, K., Middleton, A., Ambresin, G., Dowrick, C., Herrman, H., Hegarty, K., Gilchrist, G., Griffiths, F.: A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. *Journal of Affective Disorders* **148(2)**, 338–346 (2013)
9. Haken, H.: *Information and self-organization: a macroscopic approach to complex systems*. Springer (2006)
10. Jain, A.: Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*. (2009)
11. Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V., Saarela, A.: Self organization of a massive document collection. *Neural Networks, IEEE Transactions on* **11(3)**, 574–585 (2000)
12. Kuo, R., Ho, L.M., Hu, C.M.: Cluster analysis in industrial market segmentation through artificial neural network. *Computers and Industrial Engineering* **42(2)**, 391–399 (2002)
13. Kuo, R., Ho, L.M., Hu, C.M.: Integration of self-organizing feature map and k-means algorithm for market segmentation. *Computers and Operations Research* **29(11)**, 1475–1493 (2002)
14. Lasota, A., Mackey, M.C.: *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. Springer-Verlag, New York (1994)
15. Mackey, M.C.: *Time's Arrow: The Origins of Thermodynamic Behavior*. Springer Verlag, Germany (1992)
16. Mitchell, M.: *Complexity: A Guided Tour*. Oxford University Press (2009)
17. Rajaram, R., Castellani, B.: Modeling complex systems macroscopically: Case/agent-based modeling, synergetics and the continuity equation. *Complexity* (2012). DOI 10.1002/cplx.21412,2012
18. Rajaram, R., Castellani, B.: The utility of non-equilibrium statistical mechanics, specifically transport theory, for modeling cohort data. *Complexity* (2014). DOI 10.1002/cplx.21512,2014
19. Rajaram, R., Castellani, B., Gunn, J., Uprichard, E., Byrne, D., Griffiths, F.: Modeling depression and physical wellbeing longitudinally: A case-based density approach. to be submitted for review to the *American Journal of Public Health* (2015)