WILEY | Hindawi

Research Article

Improved Method for Predicting the Performance of the Physical Links in Telecommunications Access Networks

Ferenc Lilik [ID],[1] Szilvia Nagy [ID],[2] and László T. Kóczy [ID][2,3]

[1]*Department of Telecommunications, Széchenyi István University, Egyetem tér 1, Győr 9026, Hungary*
[2]*Széchenyi István University, Győr, Hungary*
[3]*Budapest University of Technology and Economics, Budapest, Hungary*

Correspondence should be addressed to Ferenc Lilik; lilikf@sze.hu

A novel approach is presented which is able to predict the available maximal data transfer rate of SHDSL connections from measured frequency dependent electrical parameters of wire pairs. Predictions are made by a fuzzy inference system. The basis of the operable and tested method will be introduced, then an improved version is shown, in which the problems derived from sampling of continuous functions of electrical parameters are eliminated by wavelet transformation. Also possibilities for simplification of the problem and a way of reducing the dimensions of the applied rule bases are presented. As the set of the measured data leads to sparse rule bases, handling of sparseness is unavoidable. Two different ways—fuzzy interpolation and various membership functions—will be introduced. The presented methods were tested by measurements in real telecommunications access networks.

1. Introduction

In telecommunications the emphasis moved from voice connections to data communications in the past decades. Despite the modern and efficient types of networks for data transmission, for example, fiber optics, huge amount of copper wire pair based networks developed for voice communications is still in use. These copper networks, especially in rural or less densely populated areas, will be in use at least in the next couple of decades.

Further use of copper networks is promoted by the development of such new methods, which aim to serve modern telecommunications services of acceptable quality on these networks. The various xDSL technologies were and are being developed to establish high speed digital connections; for example, VDSL2 technology can provide up to 200 Mbit/s data transfer rate (bit rate) on a single twisted pair [1]. Reaching this bit rate is achieved by such methods, which influences the physical circumstances of the links, like self-FEXT cancelation [2]. Thus, copper wire based access networks remain economical and applicable for providing telecommunications services for a long time.

Irrespectively of the material of transmission medium, access networks have numerous endpoints, and the performance of the access links is different in the case of each endpoint. Optimal operation, cost efficient renewal, maintenance, and forming correct business proposals and marketing strategies require information about the performance of network segments, moreover, in certain cases about the individual links. However, performance is unknown until the installation of the equipment of the given transmission technology, that is, until the formation of the real connections. Preliminary—and approximate—information about the performance can be gained by performance prediction techniques.

Performance prediction (sometimes also mentioned as performance evaluation) techniques are used in various fields of telecommunications. Lots of works deal with the examination of whole networks. Such methods can be found, for example, in [3–6]. Other works deal with simulation and soft computing based performance prediction, for example, [7, 8]. Besides the performance prediction of whole networks, prediction of the performance of individual links or channels

has a great importance; for example, Stupia et al. [9] or Brueninghaus et al. [10] deal with such solutions. Wireless performance prediction is made by Zhang and Ma [11] and performance predictions in symmetrical (copper) networks are mentioned by Magesacher et al. in [12]. Performance can be defined in different ways in the case of telecommunications networks. Some methods aim the prediction of the bit error ratio (BER) evolved in defined circumstances, for example, [13, 14]. Others, for example, Bosco et al. [15], predict the transmission power.

In this paper we present a novel method for predicting the available data transmission rate of the SHDSL connections to be installed over copper wire pairs of access networks, which makes the estimations from measured electrical parameters of the transmission medium by a fuzzy inference method, without the necessity of the installation of the technological equipment. This method—with small modifications—is suitable for making performance prognosis also for other transmission media and for other transmission technologies.

In the next section, those technical and physical parameters of the access networks that can be used for performance prediction are given together with our measurement setup. In Sections 3, 5.1, and 6.2 we summarize the theoretical background used in this contribution, namely, the fuzzy inferences, the wavelet analysis, and the KH fuzzy rule interpolation technique. In Section 4 we overview our previous results in performance prediction for wired telecommunications connections. In Section 6 using wavelet analyses and fuzzy interpolation technique we give a series of performance prediction algorithms based on solely the insertion loss values of the links of the access networks. At the end of the paper the results of the tests for SHDSL systems are shown, and we present a suggestion for a method efficiently applicable in industrial circumstances.

2. Physical Basics of Performance Prediction for Wired Links of Access Networks

As it can be read in Section 1, in this work the notation "performance" is used for the available maximal data transmission rate by SHDSL transmission. This value—as also other properties of the transmission—is influenced by the physical parameters of the transmission medium.

Studies were carried out for selecting those physical properties which really influence the bit rate. (The aim of the study was not the identification of lines out of order, but the performance evaluation of operable ones; thus the measurements were made on operable lines.) Performance primitives are given by the ITU-T recommendation referring to SHDSL technology [17]. Only two of them, loop attenuation defect and SNR (Signal to Noise Ratio) margin defect, are in connection with the layer 1 parameters of the line.

Loop attenuation defect occurs if the value of the loop attenuation is higher than a previously configured threshold. SHDSL loop attenuation is defined in page 43 of the recommendation as follows:

$$
\text{``LA} = \frac{2}{f_{\text{sym}}} \left\{ \int_0^{f_{\text{sym}}/2} 10 \log_{10} \left[\sum_{n=0}^{1} S\left(f - nf_{\text{sym}}\right) \right] df \right.
$$
$$
\left. - \int_0^{f_{\text{sym}}/2} 10 \log_{10} \left[\sum_{n=0}^{1} S\left(f - nf_{\text{sym}}\right) \left| H\left(f - nf_{\text{sym}}\right)^2 \right| \right] df \right\}, \tag{1}
$$

where f_{sym} is the symbol rate, $1/H(f)$ is the insertion loss of the loop, and $S(f)$ is the nominal transmit PSD" [17].

ITU-T also defines the SNR (in dB):

$$
\text{``SNR} = \frac{1}{M} \sum_{k=1}^{M} 10 \log_{10} \left(1 \right.
$$
$$
+ \frac{S\left(f_{\text{sym}} - f_k\right) \left| H\left(f_{\text{sym}} - f_k\right)\right|^2}{N\left(f_{\text{sym}} - f_k\right)}
$$
$$
+ \frac{S\left(f_k\right) \left| H\left(f_k\right)\right|^2}{N\left(f_k\right)} \tag{2}
$$
$$
+ \frac{S\left(2f_{\text{sym}} - f_k\right) \left| H\left(2f_{\text{sym}} - f_k\right)\right|^2}{N\left(2f_{\text{sym}} - f_k\right)}
$$
$$
\left. + \frac{S\left(f_{\text{sym}} - f_k\right) \left| H\left(f_{\text{sym}} - f_k\right)\right|^2}{N\left(f_{\text{sym}} + f_k\right)} \right),
$$

where $S(f)$ shall be the nominal far-end transmit signal power spectral density, $|H(f)|^2$ shall be the magnitude squared of the ideal loop insertion gain function, $N(f)$ shall be the injected crosstalk noise power spectral density, and f_{sym} shall be the transmit symbol rate" [17].

Although formulae (1) and (2) contain more variables, only two of them, insertion loss and noise, are real physical parameters of the line; thus, according to the recommendation, these two parameters of the line have influence on the available bit rate. This statement was verified by measurements on wire pairs from real telecommunication access networks in operation.

Measurements were performed as follows. Using operable, but actually unused wire pairs of copper cable based access networks, lines (similar to the subscriber loops) were temporarily created. Several electrical properties of the lines were studied by double-ended measurements. These properties were the noise, the far and near end crosstalk, insertion loss, return loss, line impedance, SNR, and the attenuation to crosstalk ratio. After measuring these electrical parameters, the central end of the line was connected to an SHDSL node and an SHDSL modem was installed to the other end of the line, and the available maximal bit rate was measured, granted a bit error ratio better than 10^{-7}. The results of the electrical property measurements were sorted into five clusters according to the measured bit rate ranges (range 1 to range 5, where the lowest bit rates belong to range 1 and the highest bit rates belong to range 5).

Statements of the recommendation, that is, only the insertion loss and the signal to noise ratio have influence on the performance, were verified by the measurements. Lots of the measured parameters, like, the return loss, do not have direct influence on the bit rate. Figure 1 depicts the measured return loss (RL) values. RL functions are coloured by the ranges of bit rates measured on related lines. It can be seen that although the colours seem to be grouped, there are no exclusive areas belonging to different bit rate ranges.

The role of the frequency dependent insertion loss in forming the data transmission speed is unequivocally verified by the results of the measurements. It can be seen in Figure 2 that curves of insertion loss cover areas belonging to individual bit rate ranges. The individual groups of colours are well separated; however, in some positions the bands of neighbouring ranges partially overlap.

Independently of the measured bit rates, the measured values of noise were similar to one another within each studied geographical area. Two examples for the results of noise measurements in two different areas (A and B) are shown in Figures 3 and 4. (Even though finding the reasons of similarity is not an aim of this study, according to the precursory examinations, it is caused probably by the fact that telecommunications services are similar to one another in the same areas; that is, in domestic areas analog voice and ADSL transmissions are typical; thus the noise generated by them is similar in every wire pair.) Based on the results of the measurements it can be stated that even though the noise influences the available data transmission rate, this influence is the same in the case of each line in the same area. This behavior makes possible to set up hierarchical rules which are able to evaluate the performance of the wire pairs belonging to defined areas considering only the insertion loss.

Measured data were grouped according to the available bit rates. Based on the data in these groups a system of verbal rules was formulated; their general form is as follows:

R_{0-1}: IF the form of the NOISE is similar to pattern A

 THEN the performance should be evaluated by rule base A

R_{0-2}: IF the form of the NOISE is similar to pattern B

 THEN the performance should be evaluated by rule base B

 $\vdots$

R_{0-k}: IF the form of the NOISE is similar to pattern K

 THEN the performance should be evaluated by rule base K

R_{A-1}: IF the INSERTION LOSS is huge

R_{A-2}: IF the INSERTION LOSS is large   (3)

 THEN the AVAILABLE BIT RATE belongs to range 2

 $\vdots$

R_{A-5}: IF the INSERTION LOSS is low

 THEN the AVAILABLE BIT RATE belongs to range 5

 $\vdots$

R_{K-1}: $\ldots$

 $\vdots$

Subrule base $(R_A - R_K)$, used in factual performance evaluation, is selected in the metalevel (R_0) of the hierarchical rule system described in (3).

Variables of the rule system (see (3)) are not numerical, but rather denote ranges. Moreover, these ranges are not clearly distinct; in many instances they partially cover each other. Evaluation of such rule systems by traditional mathematical tools is extremely difficult. Usage of exceedingly many equations may be needed for the exact mathematical presentation. Fuzzy logic introduced in the 1960s was developed for solving exactly such problems.

3. On Fuzzy Inferences

3.1. Basics of Fuzzy Set Theory. The usage of an exceptionally novel version of multivalued logics was suggested by Zadeh in 1965 [18]. The logic system introduced in [18] not only permits truth values of a statement differing from "YES" and "NO," but extends this possibility to the infinity. The fundamental idea of Zadeh's theoretical system is that if the phenomena of the universe can be classified into groups (sets), this classification can be numerically described and logical connections can be found between the particular phenomena, then correct and effectively handleable mathematical models of these phenomena can be produced. The operability of the system is ensured by the classifiability of the particular phenomena into special sets. The membership of the phenomena in these sets can be any value from 0 (phenomenon does not belong to the set) to 1 (phenomenon belongs absolutely to the set). Zadeh's logical system is called "fuzzy logic" or "fuzzy set theory." His paper published in 1965 lays the basics of this theory,
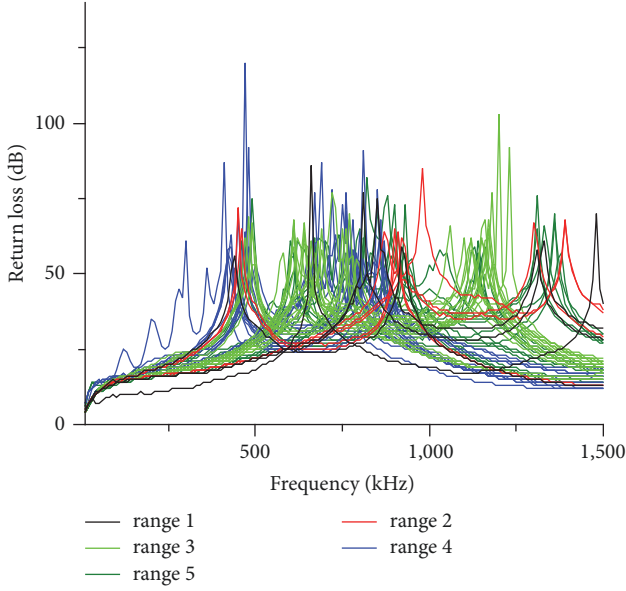
FIGURE 1: Measured return loss functions belonging to the different ranges of the bit rate.
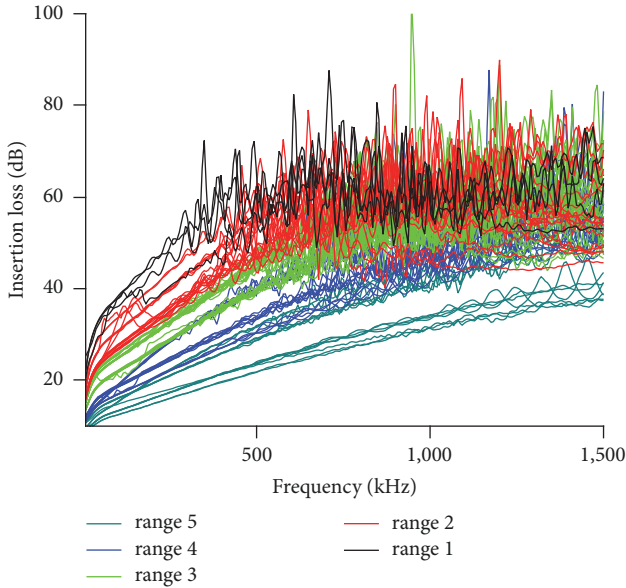

FIGURE 2: Measured insertion functions belonging to the different ranges of the bit rate.

notions of fuzzy membership value, and fuzzy membership function ($F_A(x)$—later usually denoted by $\mu_A(x)$) is defined. As an example, Figure 5 shows a possible fuzzy set of warm temperatures.

According to [18], A', that is, the complement of fuzzy set A, is defined as

$$f_{A'} = 1 - f_A, \tag{4}$$

where f_A is the membership function of fuzzy set A.
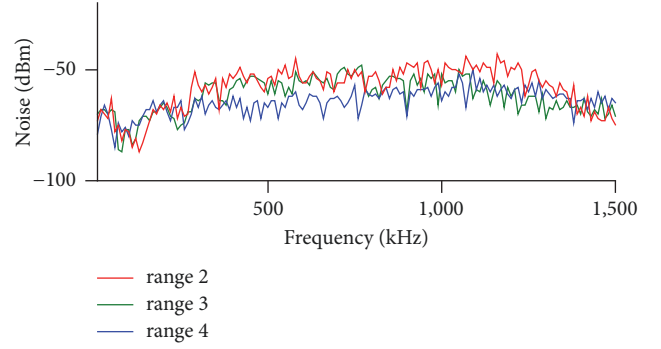

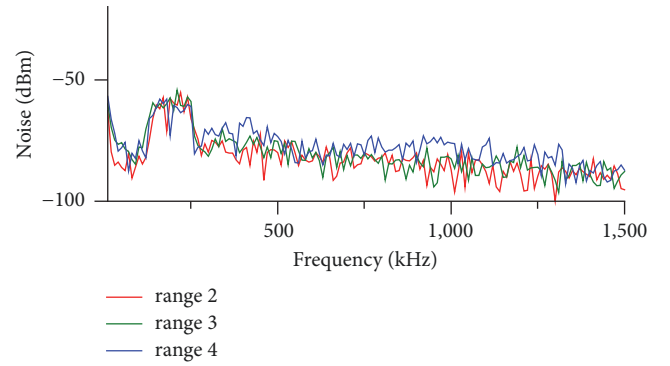FIGURE 3: Noise of lines from area A.

FIGURE 4: Noise of lines from area B.

The union (s-norm) of fuzzy sets A and B is as follows:

$$f_A \cup f_B = \max[f_A, f_B]. \tag{5}$$

The intersection (t-norm) of fuzzy sets A and B is as follows:

$$f_A \cap f_B = \min[f_A, f_B]. \tag{6}$$

Fuzzy norms can be implemented by numerous mathematical formulae; however, these formulae have to meet the conditions of the axiomatic systems of the different norms [19].

3.2. Zadeh's Fuzzy Reasoning. The basics of Zadeh's fuzzy set theory provide a good possibility for numerically solving such problems, which can be logically formulated. In his first publication in this topic, Zadeh denotes the methods appropriate for such solutions fuzzy algorithms [20]. Fuzzy algorithms are based on statements or commands in human language. Zadeh's examples are

"Set y is approximately equal to 10 if

x is approximately equal to 5"   (7)

"If x is large, increase y by several units!"   (8)

Approximately equal to 10, *approximately equal to 5*, *large*, and *several* are fuzzy sets, similar to the set of Figure 5. Also "x" and "y" are fuzzy sets; however, in most practical cases "x"
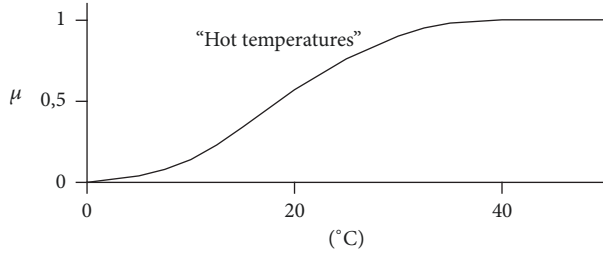
FIGURE 5: Membership values of a possible fuzzy set of hot temperatures (in °C).

is a crisp value. Using statements similar to (7) or (8), useful decision algorithms can be created, like

"If x is small, increase y by several units!

If x is large, decrease y by several units! (9)

Otherwise keep y unchanged!"

[20].

At the first time, fuzzy algorithms were used for creating fuzzy inference systems in 1973 [21]. Here, fuzzy algorithms were supplemented with logic operations, arising fuzzy IF . . . THEN rules, moreover, fuzzy rule bases, like in

R_1: IF x_1 is $A_{1,1}$ AND x_2 is $A_{2,1}$ THEN y_1 is B_1

R_2: IF x_1 is $A_{1,2}$ AND x_2 is $A_{2,2}$ THEN y_2 is B_2

$\vdots$ (10)

R_r: IF x_1 is $A_{1,r}$ AND x_2 is $A_{2,r}$ THEN y_r is B_r.

Here, $(x_1, x_2, \ldots, x_r)$ is an observation vector which can consist of fuzzy values, too, $A_{d,r}$ is a fuzzy set, where d is the number of the dimensions, r is the number of the rules, y_r is the result of the rth rule (generally also it is a fuzzy value), and B_r is the fuzzy set belonging to the output of the rth rule. The antecedent dimensions of the rules are connected to each other by AND, and the rules R_1 and R_2 by OR operator. The rule base is defined as a fuzzy relation, and the final conclusion is gained as a fuzzy composition.

"IF–THEN" fuzzy rules are used in recent control problems, too [22–24].

3.3. Mamdani's Fuzzy Inference Method. Mamdani and Assilian suggested a more simple and practicable fuzzy reasoning method in 1975 [25]. Due to its simple implementation we also use Mamdani-like approach in this work, so its operating principles are detailed here.

Similar to Zadeh's approach, the reasoning is based on two essential components. The first one is a rule base, and the other one is a quantified observation, which has to be evaluated by the rule base. The rule base contains the preliminary knowledge about the studied problem. It can be constructed from previous measurements of the system, or from experts' knowledge. The possible values of the input variables are considered as fuzzy values and thus are represented by fuzzy

sets. The outputs of the rules are also fuzzy sets (outputs of the real systems are fuzzified). It has to be mentioned here that also other approaches exist; for example, Takagi and Sugeno published a fuzzy inference system in 1985, where the rules had function type outputs [26]. The other essential component is the observation itself. The observation vector contains the values of the measurements of the examined system. Theoretically, its values are considered as fuzzy values; however, in practice, most commonly they are crisp numbers.

Basic elements of the calculations are the membership values $\mu_{d,r}$ of the singular values of the observation vector $\vec{x}$ in the fuzzy sets $A_{d,r}$ of each rule's relevant dimension. These values are in AND logic connection with each other in the same rules, where AND corresponds to the intersection set operation. In Mamdani's approach Zadeh's fuzzy intersection, namely, the "minimum", is selected to represent the logic AND. The weights w_r of the rules are calculated by

$$w_r = \min(\mu_{1,r}, \mu_{2,r}, \ldots, \mu_{d,r}),$$ (11)

and the result of the rth rule is

$$B_r^* = \min(w_r, B_r).$$ (12)

The logic OR between the rules are represented by Zadeh's s-norm; thus the final conclusion is

$$B^* = \max(B_1^*, B_2^*, \ldots, B_r^*);$$ (13)

that is, the max–min composition of the observation vector $\vec{x}$ and the rule base A^*.

Mamdani's reasoning system is illustrated in Figure 6.

In most of the cases the result of a fuzzy reasoning system is also fuzzy (Takagi-Sugeno method is an exception). Although the fuzzy result can be used by another fuzzy system, for a direct application it has to be converted to a crisp value. Several methods deal with this so-called defuzzification. Two well-known ones are Center of Gravity (14) and Center of Maxima (15).

$$y_{COG} = \frac{\sum_{i=1}^r (y_i^* \cdot w_i^*)}{\sum_{i=1}^r w_i^*}$$ (14)

$$y_{COM} = \frac{\inf M + \sup M}{2}.$$ (15)

A graphical example of the differences between the results of these two techniques can be seen in Figure 7.

4. Performance Prediction for Wired Links of Access Networks

According to the hierarchical verbal rule system described by (3) and the measured values, fuzzy rule bases were constructed. The main problem in using insertion loss as antecedent is that insertion loss is a continuous function in the frequency, which does not have a closed formula. This would mean infinite possibilities for selecting antecedent dimensions, from using insertion loss values at one or more frequency points through averages to fitting a function with
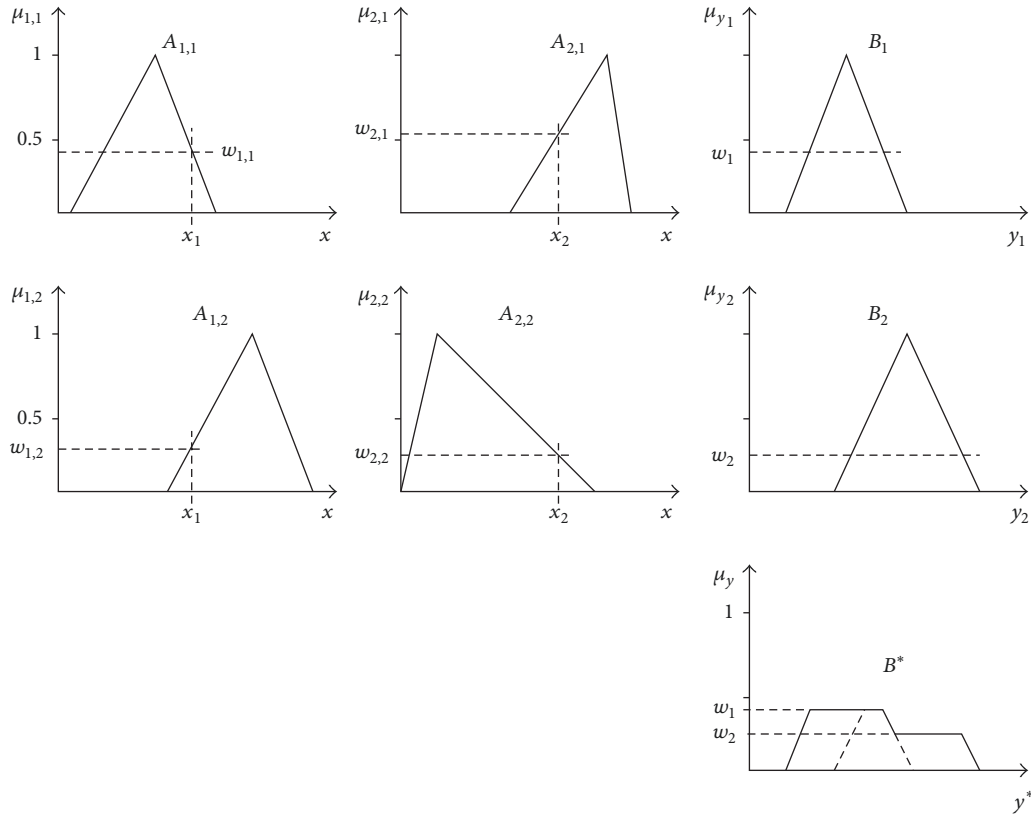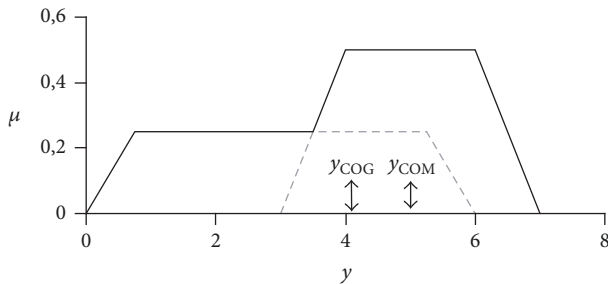
FIGURE 6: Mamdani's method.

FIGURE 7: COG and COM defuzzification of a fuzzy set (y_{COG} = 4.016, y_{COM} = 5).

some parameters to the measured values and using the fitting parameters as input vectors for the inference system.

In order to minimize the calculation, in the rule bases the characteristic points of the insertion loss were used. Based on the results of our previous research, these characteristic points in the measured 1.5 MHz wide frequency band can be at frequency points 100 kHz, 500 kHz, 750 kHz, 1000 kHz, 1250 kHz, and 1500 kHz [27]. Examinations verified that usage of more frequency points is redundant, while in case of less frequency points the effect of extreme deviations at selected frequencies can be disadvantageously dominant [28].

For the construction of the rule bases, two methods were used. In both cases the measured telecommunication access lines were separated into two parts. The training set consisted of about half the measured lines belonging to each of the performance groups. The data of the remaining lines were used for testing.

In case of the first method, such rule bases were created in which each individual rule belonged uniquely to one of the output states. The rules had six antecedent dimensions, accordingly to the measured values of the insertion loss at the six characteristic frequencies. Fuzzy sets of these rule bases were triangular with the mean value from the training set being the α-cut belonging to $\alpha = 1$ and the minimum and maximum values of the training set determining the minimum and maximum of the support of the membership function (i.e., the $\alpha = 0$ α-cut). Its graphical example can be seen in Figure 8. Later, this rule base is referred to as RB1.

Other rule bases were created by an evolutionary algorithm [29]. The rule bases were 6 dimensional also in this case; however, the number of the rules was 10 instead of 5, and the shape of the fuzzy sets was trapezoidal. A graphical example can be seen in Figure 9.

During the evaluation Mamdani's reasoning method was used, and the fuzzy results were converted into crisp values by Center of Gravity defuzzification described by (14) [30]. The rule bases were tested using the previously selected lines from operating telecommunications access networks. In case of the lines that could be evaluated the success rate of the evaluation by each rule base was better than 94%.

During the tests, two imperfections of the evaluation system had been discovered.
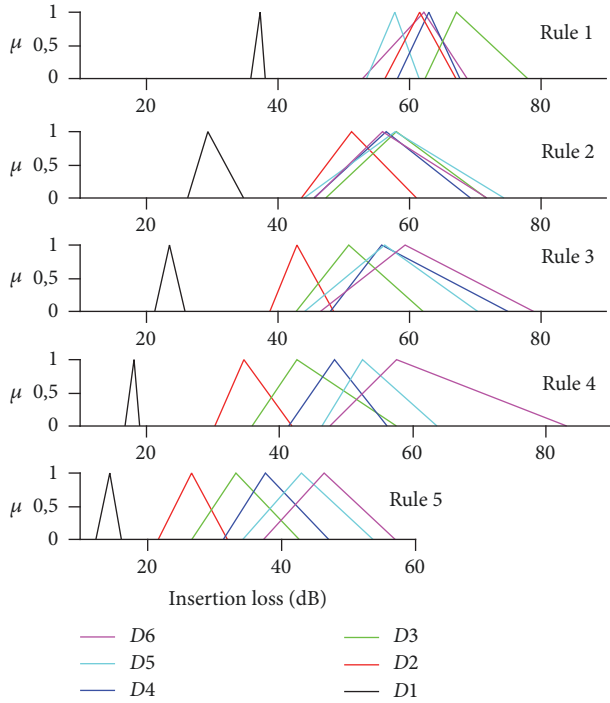
FIGURE 8: Antecedents of one of the rule bases constructed directly from measured data. Dimensions are denoted by $D1$–$D6$.
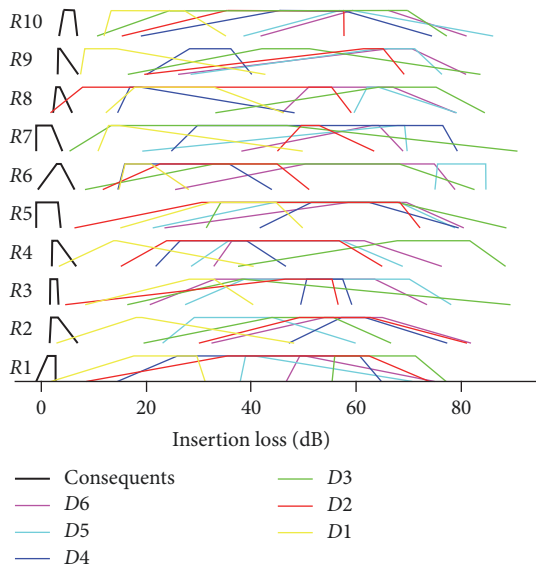

FIGURE 9: Rules 1–10 of a rule base made by evolutionary algorithm. Antecedents and consequents are trapezoidal normal (i.e., their maximum value is 1) fuzzy sets. Dimensions are denoted by $D1$–$D6$.

One of them is originated from the sparseness of the rule bases. This problem will be referred to as vertical sparseness. As the number of the lines used in rule base construction was limited, the antecedent fuzzy sets did not cover the whole input state space; thus during the construction sparse rule bases arose. Due to the sparseness of the rule bases only those

TABLE 1: An incorrect performance prediction [16].

Rule	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	Weight
$R2$	0.081	0.022	0.479	0.832	0.432	0.526	0.0222
$R3$	0	0.848	0.854	0.852	0.223	0.791	0

lines could be successfully evaluated where the measured parameters belonged to areas covered by antecedent sets. It is important that the rule base made by evolutionary algorithm— thanks to the wider supports of fuzzy sets—was able to evaluate a higher number of lines.

The other problem will be called horizontal sparseness. As the system is based on insertion loss values at very few characteristic frequencies, extreme deviations at some of these frequencies might influence the prediction of performance but not influence the bit rates, causing this horizontal type of imperfection.

The following sections of the paper discuss the solutions for these two problems and introduce an improved approach.

5. Predictions from Insertion Loss Data of the Complete Measured Frequency Domain

As a first step let us consider the problem of horizontal sparseness.

In this case the deficiency of the method was that the performance was predicted using only some discrete values of the continuous insertion loss function, even though the whole function influences the available bit rate. Clearly the method leaves out of consideration the remaining values of the insertion loss at the other frequencies; moreover it is very sensitive to point-like deviations in the measurements. Although, thanks to the properly selected frequency points, incorrect predictions occurred in minimal number during the tests; the correct results are not guaranteed. Such an incorrect prediction can be seen in Table 1.

In case of the wire pair of Table 1, the insertion loss values at the 1st dimension and its narrow surroundings is higher than in case of the lines used in rule base construction. Moreover, even though the discrepancy is actually not high, the measured value reaches the 2nd range of the rule base. This prediction is failed, as the line, according to the bit rate measurements belongs to 3rd range; that is, the deviation of the insertion loss values at dimension 1 is so small and belongs to so narrow frequency band that the line in reality can fulfill the demands of range 3 (as it can be seen in case of the other dimensions). As we have mentioned previously, the continuous function of the insertion loss offers infinite possibilities for selecting antecedent dimensions.

Our first approach that needed technical knowledge for selecting the characteristic frequencies, but no computation from the measurement values was not successful in all the cases due to measurement uncertainty and fluctuations at the characteristic frequencies. Clearly a computationally more demanding method should be used in order to get rid of the fluctuations and take into account the insertion loss in the whole frequency domain.

5.1. Wavelet Transformations. Wavelet theory is mainly used for analyzing data, for mapping the finer and rougher structure of spatial or temporal patterns. In our case both the insertion loss and the noise are in the frequency domain; however, they have finer and rougher structure, thus naturally arising the question whether the fine structure of the data influences the performance, or it can be really neglected. It is also important to study whether an average behavior represents the insertion loss and noise better than the previously selected characteristic points. The best tool for performing both of the above tasks is to use wavelet analysis on the measured data. In order to clarify the notations, we summarize the wavelet theory in the following paragraphs. For more detailed descriptions, see, for example, [31].

Discrete wavelet analysis or multiresolution analysis (MRA) represents the functions of a Hilbert space at resolution levels. These resolution levels constitute subspaces V_j of the whole Hilbert space; each rougher subspace is embedded into the finer ones; that is, $\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots$. Each of the subspaces V_j is expanded by a set of basis functions, the so-called scaling functions ϕ_{jk} with k being the shift index and j the resolution index. All the elements of all these scaling function basis sets are generated from one mother scaling function ϕ by shrinking/extending and shifting on a regular grid as

$$\phi_{jk}(x) = 2^{j/2}\phi\left(2^j x - k\right). \tag{16}$$

Note that the grid is different at each resolution level; that is, for finer resolution the grid has smaller grid distance.

At a given resolution level a function $F(x)$ can be approximated as a linear combination of the scaling functions of that resolution level,

$$F(x) \approx F^j(x) = \sum_{-\infty}^{\infty} f_{jk}\phi_{jk}(x) \tag{17}$$

with the expansion coefficients

$$f_{jk} = \int_{-\infty}^{\infty} F(x)\,\phi_{jk}(x)\,dx. \tag{18}$$

As the subspaces V_j are embedded into one another, each function $F \in V_j$ is also element of all the finer resolution level subspaces V_{j+i}. This statement is valid also for the scaling functions, thus resulting in a refinement equation

$$\phi(x) = 2^{1/2}\sum_{i=0}^{N_s} h_i\phi\left(2x - i\right). \tag{19}$$

The constants h_i are characteristic for the various basis function systems; the number N_s gives the support of the scaling functions. For the simplest scaling function type, i.e., for the Haar scaling functions, the number of the coefficients N_c is 2, and these coefficients are $h_0 = h_1 = 1$. In the case of a more complex, $N_c = 4$ scaling function, the Daubechies-4 function is a good example, its coefficients are $h_0 = 1/\sqrt{2} \cdot (1 + \sqrt{3})$, $h_1 = 1/\sqrt{2} \cdot (3 + \sqrt{3})$, $h_2 = 1/\sqrt{2} \cdot (3 - \sqrt{3})$, and $h_3 = 1/\sqrt{2} \cdot (1 - \sqrt{3})$.

The subspaces complementing V_j in V_{j+1} are the so-called wavelet subspaces or detail spaces, and their basis functions are the wavelets. Wavelets, similarly to the scaling functions, are generated from one mother wavelet ψ as

$$\psi_{jk}(x) = 2^{j/2}\psi\left(2^j x - k\right); \tag{20}$$

moreover, they are related to the scaling functions by a formula similar to the refinement equation (19)

$$\psi(x) = 2^{1/2}\sum_{i=-N_s+1}^{1} (-1)^i h_{-i+1}\phi\left(2x - i\right). \tag{21}$$

Using scaling functions as a basic resolution level approximation to a function and adding refinements to it can result in a fine resolution level approximation

$$F^j(x) = \sum_{-\infty}^{\infty} f_{0k}\phi_{0k}(x) + \sum_{i=0}^{j-1}\sum_{-\infty}^{\infty} g_{ik}\phi_{ik}(x), \tag{22}$$

where the wavelet expansion coefficients are

$$g_{ik} = \int_{-\infty}^{\infty} F(x)\,\psi_{jk}(x)\,dx. \tag{23}$$

Of course, as the fine subspace V_j is constituted from the rougher scaling function and wavelet subspaces V_{j-1} and W_{j-1} as

$$V_j = V_{j-1} \oplus W_{j-1}, \tag{24}$$

there exists a transformation, similar to the refinement equation (21) and formula (21), from the rougher level wavelets and scaling functions to the finer resolution scaling functions.

In digital data analysis we practically start from a fine resolution representation of a continuous signal (f_{jk} with a large j); this is the starting data vector (or matrix if the dimension is higher). Using the refinement equation (19) the rougher level expansion coefficients can be easily calculated from the finer resolution ones by a simple convolutional filter with filter coefficients h_i and a so-called downsampling: omitting every second element from the resulting data vector. Similar transformation produces the wavelet coefficients, with the filter constants being $(-1)^i h_{-i+1}$. Practically the transformation to the scaling function coefficients is a low-pass filtering (at about half the frequency of the highest frequency present at the starting vector), and the transformation to the wavelet coefficients is a high-pass filtering. The scaling function or low-pass coefficients describe the average behavior of the data vector, whereas the wavelet coefficients give the fine details of it.

5.2. Improved Method for Horizontal Sparseness. The problem of horizontal sparseness was solved by wavelet analysis introduced in Section 5.1. As a first step, the 1.5 MHz wide frequency range of the studied SHDSL transmission was divided into six ranges, keeping the former structure of the
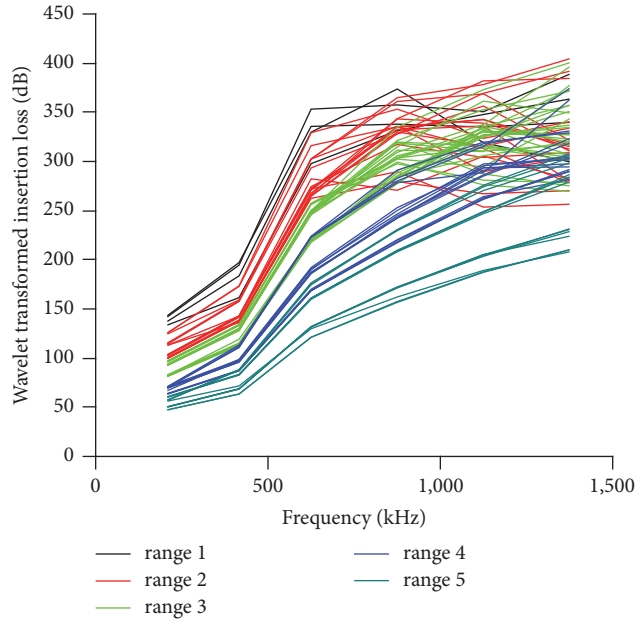
FIGURE 10: Results of the wavelet transformation of insertion loss functions shown in Figure 2.

rule bases. New rule bases were constructed using the wavelet transformed values of these six ranges. The ranges were formed as follows. The whole range was wavelet transformed into 10, than 5 values, resulting in a finer and rougher resolution representation. Due to the fact that the transmission power density is the highest in the low frequency domain, the two lowest values of the finer resolution data were used for the two lowest dimensions and the upper 4 values of the rougher resolution data for the upper four dimensions, resulting in a mixed resolution representation. Later, this type of rule bases is referred to as RBW. Figure 10 shows the wavelet transformed points. The results of the wavelet transformations are discrete values; however, for better visibility points corresponding to the same telecommunication line are connected.

According to our results, high frequency components of the wavelet transformations are not characteristic of this problem, as the available bit rate is influenced by the average heights of the insertion loss function and not by its fluctuations; thus they were not used in rule base construction. The measured insertion loss function of the line to be evaluated has to be wavelet transformed for the performance prediction. Even though this prediction seems to be made also by six discrete values, these six discrete values contain the information of the whole insertion loss function.

After doing the tests again two issues were experienced. Firstly, to achieve this result, examination of the noise was not necessary, so the hierarchical rule bases can be replaced by such a general rule base, which is constructed using wavelet transformed values of the insertion loss functions solely. Secondly, the prediction was correct in case of all wire pairs with insertion loss values belonging to the domains of the antecedent sets of the rule base. However, the problem of sparse rule bases, that is, the vertical sparseness, still existed.

6. Improvements for Handling the Sparse Rule Bases

The problem of the vertical sparseness needs other methods to overcome.

A significant degree of unsuccessful prediction is caused by the sparseness of the rule bases in case of both prediction methods introduced in Sections 4 and 5. We found two possibly efficient ways to handle this problem. One of them is the usage of fuzzy sets given by more complex membership functions, where the supports are extended to include the uncovered parts of the frequency domains in the rule antecedents. The other one is the usage of fuzzy interpolation. Both techniques were tested with the wavelet transformed functions as formerly rule bases based on wavelet transformation resulted in better predictions. The methods and the test results are described in the following.

6.1. Composite Membership Functions in Rule Antecedents. In our first method, the sparseness of the rule base in Figure 8 was canceled by using fuzzy sets described by various composite functions that had wider supports.

The first type of the composite functions consisted of four linear segments (later referred to as skirted), the second one was complete symmetric Gaussian functions (later: Gaussian), and the third one consisted of two half-Gaussian (later: semi-Gaussian) functions instead of simple triangular membership functions. The shape of these functions was formed to be similar to the shape of the original triangular antecedent sets; however, the α-cuts belonging to very small αs were much wider than the ones of the original triangular sets. All the composite rules were created from their corresponding measured and transformed minimal, maximal, and average values: $IL_{\min}$, $IL_{\max}$, and IL_{avg}.
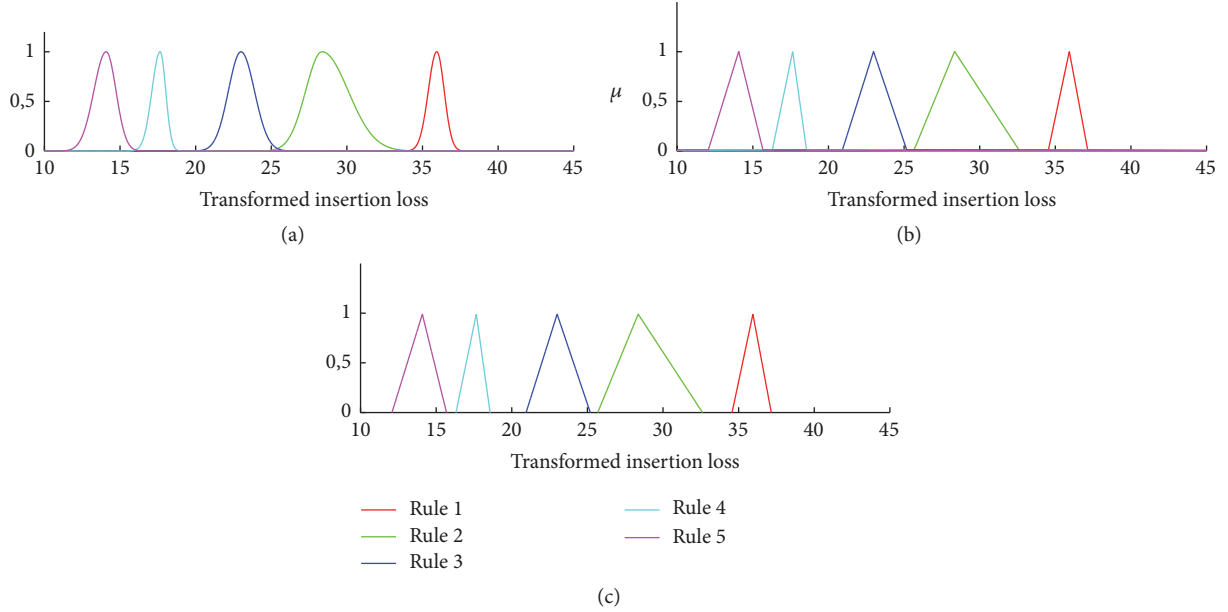
Figure 11: Compound membership functions of a rule. The original sparse rule can be seen in the diagram (c). Sparseness is canceled by the expanded supports of the composite membership functions of the two upper diagrams: (a) shows the half-Gaussians and (b) shows the skirted triangles.

For creating skirted triangular rules the average value was used as the core point, and the minimal and maximal values were used as the top of the skirts at $\alpha = 0.001$, whereas the upper and lower bounds of the support were set to $IL_{avg} - 10 \cdot (IL_{min} - IL_{avg})$ and $IL_{avg} - 10 \cdot (IL_{max} - IL_{avg})$.

In case of the Gaussian function formula

$$\mu(IL) = e^{-(IL - IL_{avg})^2 / 3(IL_{max} - IL_{min})^2} \qquad (25)$$

was used. As Gaussian functions are symmetrical but our original measurement values based triangular rules were mostly asymmetric, for a better fitting we have introduced the functions consisting of two half-Gaussians. They were created similarly to the Gaussian one, with formula

$$\mu(IL) = \begin{cases} e^{-(IL - IL_{avg})^2 / 3(IL_{avg} - IL_{min})^2} & \text{if } IL < IL_{avg}, \\ e^{-(IL - IL_{avg})^2 / 3(IL_{max} - IL_{avg})^2} & \text{if } IL > IL_{avg}. \end{cases} \qquad (26)$$

Figure 11 shows an example of a new rule constructed by half-Gaussians and skirted triangles together with the original triangular rules in Figure 11(c).

6.2. Fuzzy Interpolation. The second solution to the vertical sparseness problem was suggested by Kóczy and Hirota (KH interpolation) [33, 34]. This interpolation process creates temporary rules for the uncovered domains utilizing the nearest rule to the observation, valid only for the studied cases. More sophisticated results are provided by the stabilized KH interpolation [35]. The method creates the lower and upper bounds of the characteristic α-cuts of the corresponding fuzzy conclusion B^* by the Euclidean distance between the infimum and supremum of the characteristic α-cuts of the antecedent sets and the conclusion. Stabilized KH interpolation is given as follows:

$$\inf\{B_\alpha^*\} = \frac{\sum_{i=1}^{2n} (1/d_{\alpha L}(A^*, A_i))^k \inf\{B_{i\alpha}\}}{\sum_{i=1}^{2n} (1/d_{\alpha L}(A^*, A_i))^k},$$

$$\sup\{B_\alpha^*\} = \frac{\sum_{i=1}^{2n} (1/d_{\alpha U}(A^*, A_i))^k \inf\{B_{i\alpha}\}}{\sum_{i=1}^{2n} (1/d_{\alpha U}(A^*, A_i))^k}, \qquad (27)$$

where i denotes the number of the rules, k denotes the number of the dimensions (variables), A^* denotes the observation, A_i denotes the antecedent sets in rule i, $d_{\alpha L}(A^*, A_i)$ and $d_{\alpha U}(A^*, A_i)$ denote the lower and upper bounds of the distance between the α-cuts of observation and the antecedents, and B^* stands for the corresponding fuzzy conclusion [36].

6.3. Reduction of the Number of the Dimensions. The method described in Section 4 was based on a rule base with 6 dimensions. Its reason was that the continuous functions of physical line parameters were represented by some of their discrete values at previously selected frequencies. However, later, the information of the whole insertion loss functions was taken into consideration by the usage of the wavelet transformation method described in Section 5. Based on the results of our examinations, the reduction of the number of dimensions seemed to be possible. In order to find the optimal number of the input variables (dimensions), the insertion loss functions were wavelet transformed to different resolution levels by Haar and Daubechies wavelets, and according to these transformations, different rule bases were constructed.

TABLE 2: Techniques combined for constructing new type rule bases.

Function shape/interpolation	Number of points (Haar)	Number of points (Daubechies)
Skirted	8	4
Semi-Gaussian	4	3
Gaussian	2	2
Triangular + stabilized KH interp.	1	1

The number of the resulting points after the wavelet transformations is different in order to cover approximately the same frequency domain in the two cases.

In the calculations of the next section we have applied transformations to the starting sampled insertion loss functions down to 1, 2, 4, or 8 low-pass coefficients for the Haar basis set and to 1, 2, 3, or 4 low-pass coefficients in the case of the Daubechies 4 functions. We also kept the last high-pass coefficients (that characterize the largest scale behavior of the original function). These rule bases have the shorthand notation DB for Daubechies 4 and H for Haar wavelets and L2, L4, and so on for the number of remaining points after low-pass filtering or HL1 if the wavelet transform resulted in one high-pass (wavelet) and one low-pass (scaling function) coefficient.

6.4. Results of the Improved Predicting Method. To create a sufficiently precise and efficient system, the methods described in Sections 6.1–6.3 were combined. Symmetrical Gaussian and asymmetrical skirted triangle and semi-Gaussian rule bases were constructed from data arising from various resolution level wavelet transformations of the insertion loss functions. Also the original triangular (sparse) rule base was improved by stabilized KH interpolation described by (27).

Another modification of the original method is that certain labels referring to the bit rate range are used in the outputs of the rules instead of consequent fuzzy sets (it is possible because each rule belongs to one and only one output state), and the final conclusion is the value belonging the label of the rule with the highest firing rate [30].

Combining the techniques from Table 2, 43 different rule bases were created. These rule bases were used in testing of 65 lines measured in five different telecommunications access networks in different circumstances (different areas: industrial, commercial, and residential at different times). The results of the tests were sorted into 3 clusters. Cluster *Correct* contained those predicted results which were equal to the measured ones. Predicted results, which ranked the evaluated line into the lower neighbour of the correct range of the bit rate, formed the cluster *Acceptable*. Other incorrect results belonged to the cluster *Incorrect*. The results classified in this way can be seen in Table 3.

According to Table 3, the best predictions were made by rule bases numbers 1 and 2. Both gave the same results for all the tested lines. The main reason for this equivalence is the very similar shapes of the fuzzy sets used in the rule bases: both of them were constructed from data of the 2-pointed wavelet transformation of the insertion loss functions using Haar wavelets. Even though the same results were gained by the first 2 rule bases, the usage of rule base number 1 is more effective, due to its easier applicability. The

method with the most accurate theoretical background, the interpolated sparse rule base constructed by Daubechies 4 wavelet transform, is still very effective. This is the rule base with 6 antecedent dimensions described in Section 5. The huge difference in the success rate between methods numbers 10 and 28 is caused by an incorrect normalization. As the whole input space is not known; thus the normalizing factor can not be correctly defined.

An interesting result is that the symmetric membership functions perform much worse than their asymmetric counterparts, which fit the measured values of the training set not the test set.

7. Conclusions

After a summary about the previously developed fuzzy inference based access network link performance prediction techniques we gave a series of improvements to the methods. The first improvement is using large-scale (averaged) insertion loss behavior, that is, wavelet transformed values instead of single measured values at characteristic frequencies, which makes the previous methods more robust to local insertion loss fluctuations.

The second improvement of the sparse rule bases is applying either membership functions with extended (widened) support (skirted triangles or a composite of two half-Gaussians) or stabilized KH fuzzy interpolation technique. Both types of methods make the previously not evaluable test lines evaluable, with correct results for almost all the testing lines. Interestingly very rough resolution level wavelet transforms of the insertion loss function (transformation down to 2 points) give the best results with rules having extended supports; however, the interpolation gives almost as good results with 6 antecedent dimensions. The worst results, that is, the highest number of not acceptable classifications, use Haar wavelets, large number of antecedent dimensions, and rules with extended support.

Recently the method is used for image classification purposes in mechanical and medical pictures.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

TABLE 3: Results of the tests of the different rule bases.

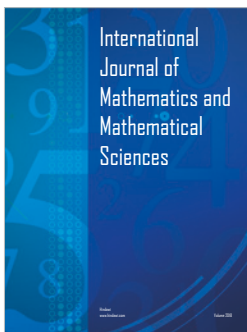| Number	Rule base	Correct	Acceptable	Incorrect
1	Skirted – H – L2	57	8	0
2	Half-Gaussian – H – L2	57	8	0
3	RBW DB with interpolation	57	7	1
4	Skirted – H – HL1	57	6	2
5	Half-Gaussian – H – HL1	57	6	2
6	Skirted – H – L4	54	11	0
7	Half-Gaussian – H – L4	54	11	0
8	Skirted – H – L8	54	4	7
9	Half-Gaussian – H – L8	54	4	7
10	RB1 with interpolation	53	12	0
11	Skirted – DB – HL1	52	12	1
12	Skirted – DB – L2	52	12	1
13	Half-Gaussian – DB – HL1	52	12	1
14	Half-Gaussian – DB – L2	52	12	1
15	Skirted – DB – L3	51	13	1
16	Half-Gaussian – DB – L3	51	13	1
17	Interpolated – DB – L3	50	15	0
18	Skirted – DB – L1	50	14	1
19	Half-Gaussian – DB – L1	50	14	1
20	Gaussian – DB – L3	49	16	0
21	Interpolated – H – L8	49	16	0
22	Interpolated – DB – HL1	49	12	4
23	Gaussian – H – HL1	48	17	0
24	Gaussian – DB – L4	48	17	0
25	Skirted – DB – L4	48	16	1
26	Half-Gaussian – DB – L4	48	16	1
27	Interpolated – DB – L2	48	16	1
28	RB1 with normalized interpolation	46	19	0
29	RBW H with interpolation	46	19	0
30	Skirted – H – L1	45	20	0
31	Half-Gaussian – H – L1	45	20	0
32	Gaussian – DB – L2	45	20	0
33	Gaussian – H – L1	44	21	0
34	Gaussian – H – L2	44	21	0
35	Interpolated – H – L4	44	21	0
36	Gaussian – H – L4	43	19	3
37	Gaussian – H – L8	43	15	7
38	Interpolated – DB – HL1	42	19	4
39	Gaussian – DB – HL1	40	23	2
40	Gaussian – DB – L1	39	24	2
41	Interpolated – H – L2	37	28	0
42	Interpolated – H – HL1	36	29	0
43	Interpolated – H – HL1	24	41	0

In column *Rule base* the first capital (H or DB) stands for the wavelet transformation method (Haar or Daubechies), the second 1 or 2 capitals (L or HL) stand for the used components of the results of the wavelet transformations (high or high and low), and the number at the last position refers to the points resulted by the wavelet transformations.

References

[1] ITU: Very high speed digital subscriber line transceivers 2 (VDSL2), Technical Recommendation G.993.2, ITU Std., 2011.

[2] ITU: Self-FEXT cancellation (vectoring) for use with VDSL2 transceivers, Technical Recommendation G.993.5, ITU Std., 2010.

[3] D. Guo and X. Wang, "Bayesian inference of network loss characteristics with applications to TCP performance prediction," in *Proceedings of the IEEE Workshop on Statistical Signal Processing, SSP 2003*, pp. 530–533, USA, October 2003.

[4] R. Pan, B. Prabhakar, K. Psounis, and D. Wischik, "SHRiNK: A method for scaleable performance prediction and efficient network simulation," in *Proceedings of the IEEE INFOCOM '03. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 1943–1953, San Francisco, CA, USA, 2003.

[5] B.-S. Chen, S.-C. Peng, and K.-C. Wang, "Traffic modeling, prediction, and congestion control for high-speed networks: A fuzzy AR approach," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 491–508, 2000.

[6] L. Muka and I. Derka, "Simulation Performance Prediction in Clouds," in *Proceedings of the 9th International Symposium on Applied Informatics and Related Areas - AIS2014*, pp. 142–147, 2014.

[7] L. Muka and I. Derka, "Evaluation and improvement of parallel discrete event simulation performance predictions: A rough-set-based approach," *Acta Polytechnica Hungarica*, vol. 13, no. 6, pp. 125–145, 2016.

[8] G. Lencse, I. Derka, and L. Muka, "Towards the efficient simulation of telecommunication systems in heterogeneous distributed execution environments," in *Proceedings of the 36th International Conference on Telecommunications and Signal Processing, TSP '13*, pp. 304–310, Italy, July 2013.

[9] I. Stupia, F. Giannetti, V. Lottici, and L. Vandendorpe, "A novel link performance prediction method for coded MIMO-OFDM systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference, WCNC '09*, Hungary, April 2009.

[10] K. Brueninghaus, T. Salzer, D. Astely et al., "Link Performance Models for System Level Simulations of Broadband Radio Access Systems," in *Proceedings of the IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications '05*, pp. 2306–2311, Berlin, Germany, 2005.

[11] J. Zhang and X. Ma, "Broadcast performance analysis of IEEE 802.11 networks under fading channels," in *Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS '13)*, pp. 19–21, Toronto, 2013.

[12] T. Magesacher, D. Statovci, T. Nordström, and E. Riegler, "Performance analysis of vectored wireline systems embracing channel uncertainty," in *Proceedings of the IEEE International Conference on Communications, ICC '13*, pp. 3986–3990, Hungary, June 2013.

[13] G. Kaddoum, F. Gagnon, P. Chargé, and D. Roviras, "A generalized BER prediction method for differential chaos shift keying system through different communication channels," *Wireless Personal Communications*, vol. 64, no. 2, pp. 425–437, 2012.

[14] K. Sayana and J. Zhuang, "Link Performance Abstraction based on Mean Mutual Information per Bit (MMIB) of the LLR Channel," IEEE 802.16 Broadcast Wireless Access Working Group, 2007.

[15] G. Bosco, A. Carena, R. Cigliutti, V. Curri, P. Poggiolini, and F. Forghieri, "Performance prediction for WDM PM-QPSK transmission over uncompensated links," in *Proceedings of the 2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference, OFC/NFOEC 2011*, usa, March 2011.

[16] F. Lilik, S. Nagy, and L. T. Kóczy, "Wavelet based fuzzy rule bases in pre-qualification of access networks' wire pairs," in *Proceedings of the 12th IEEE AFRICON International Conference, AFRICON 2015*, Ethiopia, September 2015.

[17] ITU: Single-pair high-speed digital subscriber line (SHDSL) transceivers, Technical Recommendation G.993.2, ITU Std., 2003.

[18] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[19] L. T. Kóczy and D. Tikk, "Fuzzy rendszerek (Fuzzy systems, in Hungarian)," *Typotex*, 2000.

[20] L. A. Zadeh, "Fuzzy algorithms," *Information and Control*, vol. 12, no. 2, pp. 94–102, 1968.

[21] L. A. Zadeh, "Outline of new approach to the analysis of complex systems and decision processes," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, pp. 28–44, 1973.

[22] X.-H. Chang, J. Xiong, Z.-M. Li, and J. H. Park, "Quantized Static Output Feedback Control For Discrete-Time Systems," *IEEE Transactions on Industrial Informatics*, 2017.

[23] X. Chang and Y. Wang, "Peak-to-Peak Filtering for Networked Nonlinear DC Motor Systems with Quantization," *IEEE Transactions on Industrial Informatics*, 2018.

[24] X. Xie, D. Yue, H. Zhang, and C. Peng, "Control synthesis of discrete-time T-S Fuzzy systems: reducing the conservatism whilst alleviating the computational burden," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2480–2491, 2017.

[25] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, 1975.

[26] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[27] F. Lilik and J. Botzheim, "Fuzzy based Prequalification Methods for EoSHDSL Technology," *Acta Technica Jaurinensis*, vol. 4, no. 1, pp. 135–144, 2011.

[28] F. Lilik, S. Nagy, and L. T. Kóczy, "Examination of Characteristic Frequencies of Insertion Loss in Performance Evaluation of Access Networks' Wire Pairs," *Acta Technica Jaurinensis*, vol. 8, no. 3, p. 267, 2015.

[29] K. Balázs and L. T. Kóczy, "Constructing dense fuzzy systems by adaptive scheduling of optimization algorithms," *Applied and Computational Mathematics*, vol. 11, no. 1, pp. 81–101, 2012.

[30] F. Lilik and L. T. Kóczy, "Performance evaluation of wire pairs in telecommunications networks by fuzzy and evolutionary models," in *Proceedings of the IEEE AFRICON 2013*, Mauritius, September 2013.

[31] I. Daubechies, "Ten Lectures on Wavelets," in *CBMS-NSF regional conference series in applied mathematics 61*, SIAM, Philadelphia, 1992.

[32] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme - Erste Mitteilung," *Mathematische Annalen*, vol. 69, no. 3, pp. 331–371, 1910.

[33] L. T. Kóczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *International Journal of Approximate Reasoning*, vol. 9, no. 3, pp. 197–225, 1993.

[34] L. T. Kóczy and K. Hirota, "Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases," *Information Sciences*, vol. 71, no. 1-2, pp. 169–201, 1993.

[35] D. Tikk, I. Joó, L. T. Kóczy, P. Várlaki, B. Moser, and T. D. Gedeon, "Stability of interpolative fuzzy KH controllers," *Fuzzy Sets and Systems*, vol. 125, no. 1, pp. 105–119, 2002.

[36] K. Balázs and L. T. Kóczy, "Hierarchical-interpolative fuzzy system construction by genetic and bacterial memetic programming approaches," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, no. 2, pp. 105–131, 2012.