ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 10, April 2014

Induction of fuzzy rules based on scattered data modeling and multidimensional interpolation: A novel approach

D. S. Sfiris
Department of Civil Engineering, Democritus University of Thrace, 67100 Xanthi, Greece

Abstract—this paper describes a novel approach for inducing fuzzy rules from numerical data. Multidimensional interpolation is employed to construct an interpolation grid, fit to the scattered data. This enables us to automatically obtain a complete fuzzy rule base by associating a fuzzy rule to each point of the grid. Our proposed approach is based on singular spectrum analysis to efficiently reduce noise and improve smoothness of interpolation. Finally, the derived fuzzy rules define a rule-based fuzzy inference system. An application example is provided where it is shown that the estimations are improved.

***Index Terms*—rule based inference systems, data mining, fuzzy inference, missing data, singular spectrum analysis.**

I. INTRODUCTION

Fuzzy rule models have received special attention in the literature [1]. Information is gathered by domain experts and expressed in the form of fuzzy rules, which can then be translated in order to create a fuzzy model. However, there are many cases where there are no available experts or it is difficult to explicitly state expert knowledge in the form of fuzzy rules. In such cases, inducing fuzzy rules from available data is required. Various attempts have been proposed to automatically extract fuzzy rules from experimental input/output data. Among these, learning methods based on clustering and recognition of patterns in the data [2]. Wang & Mendel's method [3] and an extension of it by Casillas *et al.* [4] are known iterative procedures for generating fuzzy rules by learning from examples. Based on Wang & Mendel's method, Hong & Lee [5] proposed a learning approach which is able to automatically derive membership functions and fuzzy if-then rules from a set of given training examples. The main restrictions in the use of clustering and pattern-based algorithms include: (1) the presence of gaps in the data which may degrade the quality of estimates, and (2) data grouping according to certain assumptions which may be unrealistic. The last restriction is encountered especially when it is difficult to discern patterns in the data. In this paper, we propose a new perspective to induce fuzzy rules from numerical data, where no assumptions are made about patterns in the sample space. In our approach, we focus on scattered data modeling and multidimensional interpolation [6]. Furthermore, we apply singular spectrum analysis [7] to cope with missing or incomplete data. Finally, we give detailed instructions for applying this perspective by using as an example a linear interpolation based on the k-nearest neighbors (k-NNS). The rest of this paper is organized as follows. In Section II, we present the problem of inducing fuzzy rules from scattered data. In Section III, we present our approach to fuzzy rule induction. Section IV demonstrates an application example of an expert system which estimates land prices. We conclude by summarizing the new concepts and discussing our results.

II. PROBLEM FORMULATION

A point cloud is a set of points in space describing the relationship between the dependent variable and one or more independent variables. Let us assume that there exists a set of n input-output training examples which constitute a point cloud, that is

$$\left(x_1^{(1)},\ldots,x_k^{(1)},z^{(1)}\right),\left(x_1^{(2)},\ldots,x_k^{(2)},z^{(2)}\right),\ldots,\left(x_1^{(n)},\ldots,x_k^{(n)},z^{(n)}\right) \quad (1)$$

where $x_j^{(i)}, j=1\ldots k$ are input values, and $z^{(i)}$ is the output value of the i th training example. Fuzzy rules can be derived from a given point cloud by attributing a fuzzy set (e.g. triangular, Gaussian, etc.) to each point of the cloud. This approach makes it feasible to derive fuzzy rules R^l of the form

$$R^l : \text{IF } x_1^l \text{ is } F_{x_1}^l \text{ and} \ldots \text{and } x_k^l \text{ is } F_{x_k}^l, \text{ THEN } z \text{ is } G^l \quad (2)$$

where F_i^l and G^l are fuzzy sets, $x=\left(x_1,\ldots,x_n\right)^{\mathrm{T}} \in U$ and $z \in V$ are input and output linguistic variables, respectively, and $l=1,2,\ldots,n$. In this way, a total of n fuzzy rules correspond to the n examples. A problem with this approach is that unless the sample data space is complete, fuzzy rules will not overlap appropriately. A necessary condition in forming fuzzy rules is that for every $x_i \in X_i$, there is at least one antecedent of a fuzzy rule, which is true over x_i at least to a fixed degree $\alpha > 0$.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 10, April 2014

Formally, if

$$\forall x_i \in X_i, i=1\ldots k, \exists \left(l \in \{1\ldots n\}\right): F_i^l\left(x_i\right) \geq \alpha > 0$$

then $\{F_i^l\}(l=1\ldots n)$ form an α -cover of X_i.

Wang and Mendel [3] proposed a five-step partitioning method which limits the number of rules based on the number of training pairs. Following Wang and Mendel's approach, Hong & Lee [5] proposed a heuristic method based on clustering of the outputs of the training data and then mapping them into linguistic variable terms. Other excellent works on the topic can be found in [8] and [9]. A key idea of these works is to approximate an unknown non-linear function to derive fuzzy systems including membership functions characterizing linguistic terms in fuzzy rules. However, it should be noted that such approaches do not always guarantee stability due to presence of gaps in real data, lack of discernible structure, and possible errors in interpolation due to missing data. These problems have motivated us to develop a new fuzzy rule induction approach in which no assumptions concerning the spatial distribution of the data are made. The approach involves three stages: (1) construction of an initial approximation grid of a desired resolution to fit the scattered data; (2) set up of the final *reconstructed approximation grid*, by filtering noise and removing inconsistencies; and (3) deriving fuzzy rules corresponding to the reconstructed grid nodes. In the following section, we describe each stage in detail.

III. THE PROPOSED RULE INDUCTION APPROACH

Our proposed approach deals with the induction of fuzzy rules from a properly constructed approximation grid which is fit to the scattered data. In the first stage, we deal with approximation of a real valued function $f: R^d \to R$ on a finite set of distinct input points $X = \{x_1, \ldots, x_n\} \subset R^d$, where $d \geq 1$ is the dimension of the input space R^d . This allows us to derive the fuzzy rules indirectly from the grid points instead of the cloud points. It also allows us to denoise the point cloud and deal effectively with missing values or erroneous values. We now provide the required steps from multidimensional interpolation up to the final acquisition of the fuzzy rules.

A. *Fitting an initial interpolation grid*

Given a point cloud, such as

$$\left(x_1^{(1)},\ldots,x_k^{(1)},z^{(1)}\right),\left(x_1^{(2)},\ldots,x_k^{(2)},z^{(2)}\right),\ldots,\left(x_1^{(n)},\ldots,x_k^{(n)},z^{(n)}\right)$$

we build an interpolation grid fit by choosing a grid resolution and any interpolation technique such as kriging, polynomial regression, linear interpolations, nearest neighborhood approximation, and others [10]. For sake of example, here we use linear interpolation based on the k-nearest neighbors (k-NNS) to compute the coordinates of each grid node. Either linear interpolation or extrapolation may be required, considering that each grid node belongs to a local plane. In case of many dimensions, a $m-1$ dimensional space more is defined by the nearest neighboring to the node cloud points. In three dimensions, three nearest cloud points neighboring to each grid node are sufficient to define a plane $f(x,y) = \mathrm{A}x + \mathrm{B}y + \mathrm{C}$ which can be computed algebraically by solving for A, B and C. Note that the three cloud points which define a local plane should not be collinear or near-collinear. Thus it is necessary to perform a co linearity test before choosing neighboring cloud points. Depending on the chosen interpolation method, a more or less accurate initial fit of the grid to the scattered data is obtained. However, it should be noted that when there is noise or errors in the data, an optimal fit has no sense as it gives a false impression of correctness, while, actually, this kind of over fitting to erroneous data may become a source of errors by itself. The next step is to make some preprocessing to denoise the data. This task is accomplished by employing Singular Spectrum Analysis (SSA) [11] on the initial interpolation grid. The next step explains how to denoise the grid by employing SSA filtering.

B. *Filtering noise by means of SSA*

The key idea is to consider the grid lines to be discrete time series, permitting us to apply singular spectrum analysis on each of these time series separately. One dimensional time series SSA is applied to each series of nodes of the grid for all grid dimensions. Each grid line is then reconstructed from the decomposed principal components, particularly the first few ones, the rest of them considered to be noise. How many principal components to use is problem dependent, according to the quality of the data and the existence or not of noise. The procedure is repeated on all grid sections. Below we provide a brief description of the methodology of the SSA technique. In doing so, we mainly follow Golyandina *et al* [11].

Let $F = \left(f_0, f_1, \ldots f_{N-1}\right)$ be a time series of length N and L be an integer called embedding parameter.

1) *Construction of the trajectory matrix*

$$\mathrm{X} = \left(f_{i+j-2}\right)_{i,j=1}^{L,K} = \left[X_1 : \ldots : X_K\right]$$

where $K = N - L + 1$, $j = 1, \ldots K$.

2) *Computation of the eigen values and eigenvectors of the*

X *is a sum of rank-one biorthogonal matrices*

$\mathrm{X}_i\left(i=1,\ldots,d\right)$ *where* $d\left(d \leq L\right)$ *is the number of nonzero singular values of* X.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 10, April 2014

3) Split of the set of indices $I = \{1, \dots, d\}$ into several groups $I_1, \dots, I_m$ and sum of the matrices X_i within each group. The result of the step is the representation $X = \sum_{k=1}^{m} X_{I_k}$, where $X_{I_k} = \sum_{i \in I_k} X_i$.

4) Average over the diagonals $i + j$ of the matrices X_{I_k}. This gives us a decomposition of the original series F into a sum of series $f_n = \sum_{k=1}^{m} f_n^{(k)}, n = 0, \dots, N-1$, where for each k the $f_n^{(k)}$ is the result of diagonal averaging of the matrix X_{I_k}

The result of SSA is a decomposition of the initial series into a number of additive series or principal components. We can then reconstruct the original series using only a few of the principal components thus removing the noise. Usually, three or four principal components are adequate for most applications. In this way, the initial interpolation grid is smoothed by applying SSA on each row and column of the grid, without assuming any functional dependence between inputs – output variables. The next step shows how to compensate for possible inconsistencies occurred during this process.

C. Setting up the final grid by means of averaging

Each grid node belongs to various intersecting grid lines. Due to the separate SSA procedures along the different dimensions, the calculated coordinates of each grid node derived from various grid lines do not coincide. Since discontinuities are not allowed, we average the results. This leads us to the final reconstructed interpolation grid. The disadvantage of having to compensate for discontinuities is a favorable trade-off to actually reduce the complexity of fitting and filtering by a degree of magnitude. For example in case of three dimensions a $n \times m$ fit problem is reduced into a $n + m$ problem. More iterations of SSA on the result grid can always be applied to further improve the total mean squared error.

D. Fixing erroneous grid nodes

There are reasons that some grid nodes have to be fixed. As it happens in statistics, some data may be unrealistic and completely unacceptable due to errors in the data or some other factor. Such data would not lead to a fair interpolation grid and therefore erroneous grid nodes have to be removed. A modified Singular Spectrum Analysis (SSAM) proposed by Schoellhamer [7], which is SSA with missing data, can easily fill in the gaps and process such missing nodes. Eigen values and eigenvectors are computed for the lagged autocorrelation matrix. The improved SSAM approach when there are missing data in the grid is as follows.

1) Computation of the lagged autocorrelation by ignoring and pair of data points with a missing value

$$\bar{c}_j = \frac{1}{N_l} \sum_{l \le N-j} \bar{x}_l \bar{x}_{l+j}, \quad \textit{Where} \quad 0 \le j \le M-1 \quad \textit{for} \quad N_l$$

pairs with no missing data.

2) Computation of eigenvalues and eigenvectors for the lagged autocorrelation matrix as with SSA.

3) Computation of the k th principal component ignoring missing data points

$$\bar{a}_i^k = \frac{M}{N} \sum_{l \le M} \bar{x}_{i+l} \bar{E}_l^k, \quad 0 \le i \le N - M, \textit{ for } N_l x_{i+1} \textit{ with no}$$

missing data.

4) If $N_l < fM$, where $0 \le f \le 1$ is a specified fraction of allowable missing data points within window size M, then $\bar{a}_i^k$ is assigned a missing value.

Reconstructed components are calculated as with SSA.

E. Induction of fuzzy rules based on the reconstructed approximation grid

The next step is to attach fuzzy sets centered to each grid node. This ensures uniform overlapping of fuzzy sets. The membership function representing the fuzzy sets may be defined as a triangular fuzzy number whose support is extending to the length of two adjacent grid nodes (Fig. 1) or more (Fig. 2), up to the total length of the sample space (Fig. 3).

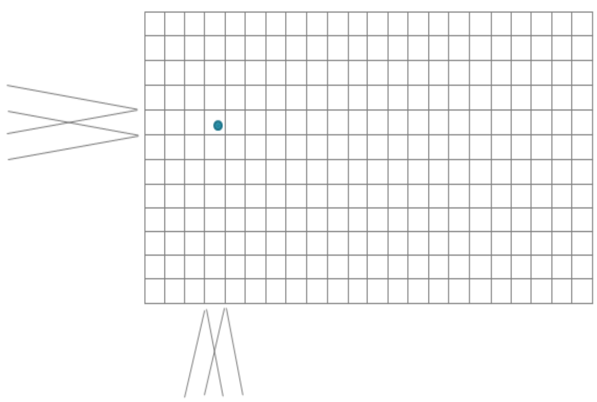

Fig. 1 Fuzzy sets with narrow support extending to the length of two adjacent grid nodes

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 10, April 2014

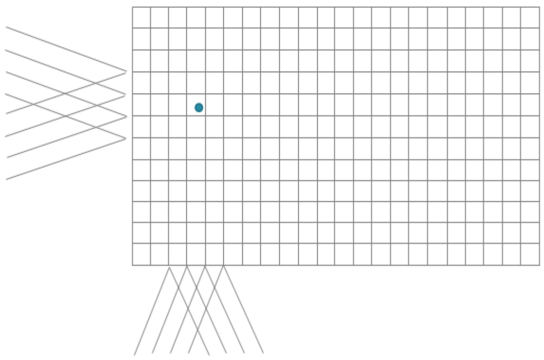

Fig. 2 Fuzzy sets with broader support extending to the length of further adjacent grid nodes

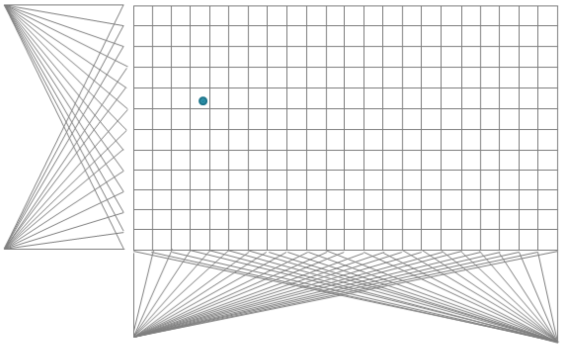

Fig. 3 Fuzzy sets with full support extending to the total length of the sample space

The broader the support of fuzzy sets, the bigger the influences of the adjacent nodes on each grid node become.

A. Defining a fuzzy-inference system

The above fuzzy sets comprise the antecedents of fuzzy rules of a fuzzy inference system. In three dimensions for example, for any arbitrary (x,y) location, the corresponding z coordinate can be determined by fuzzy reasoning after firing a number of fuzzy rules. Because of overlapping, a number of fuzzy rules will always be fired depending on the partial matching of the (x,y) location and the fuzzy sets attached to the grid nodes. The Takagi-Sugeno (T-S) fuzzy model (Fig. 4), proposed in Takagi and Sugeno [12], is computationally efficient and guarantees the continuity of the output surface. Furthermore, it has been shown that T-S fuzzy models are capable of approximating any smooth non-linear functions to any degree of accuracy [13,14].

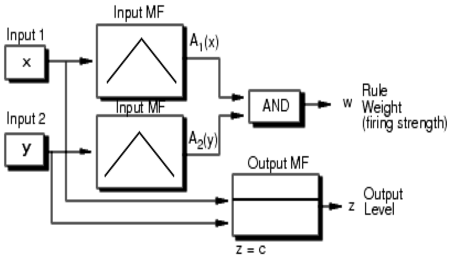

Fig. 4 Block diagram of Takagi-Sugeno model

To obtain representative estimates of model outputs through fuzzy composition, the firing strength w_i of each fuzzy rule is computed by the conjunction $w_i = \mu_{A_{i1}}(x_1) \wedge \mu_{A_{i2}}(x_2) \wedge \cdots \wedge \mu_{A_{ir}}(x_r)$, where $\wedge$ is a suitable t-norm? We refer to [15] for a detailed description of T-S fuzzy inference. The next section demonstrates an application example to clarify our proposed framework.

IV. AN APPLICATION EXAMPLE

Let us consider a real estate agency with data of land prices acquired from contracts of past transactions, adjusted for inflation. Suppose that

$$S = \left\{ \left(x^{(1)}, y^{(1)}, z^{(1)}\right), \left(x^{(2)}, y^{(2)}, z^{(2)}\right), \ldots, \left(x^{(50)}, y^{(50)}, z^{(50)}\right) \right\}$$

is a set of 50 price examples where $\left(x^{(i)}, y^{(i)}\right)$ are the coordinates of the location of a particular piece of land and $z^{(i)}$ is its price per square meter. Each particular example cannot be considered to be representative of a correct estimate of land price in the neighboring area, since it is usually a product of negotiations. Table 1 displays the scattered data cloud along with the estimated prices after our methodology was applied. As it can be seen, the given point cloud has considerable variation. Fig. 5 displays the original point cloud. First, a 100x100 approximation grid was initially constructed as described in Section III-A. Then, SSA was applied on all grid lines to decompose each one of them into ten principal components. Of all the ten decomposed components, only the first three were used to reconstruct the interpolation grid, thus considerably denoising the data. Fig. 6 shows a detail of the application of SSA on one of the grid lines. In Fig. 7, five points have been removed from the series, to show how little any missing values would affect the result. Fig. 8 shows the final reconstructed 100x100 approximation grid. Finally, we tested zero-order Takagi-Sugeno model and t-norm as a conjunction of the fuzzy rules. The result of the expert system was found to be quite satisfactory in approximating the real values. The Mean Percentage Error (MPE) was found to be

$$\mathrm{MPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{f^{(i)} - z^{(i)}}{z^{(i)}} = 0.0391.$$

A distribution of the expert system results is shown in a contour plot in Fig. 8.

Table 1. Initial price data and estimates of the expert system

ID	$x^{(i)}$	$y^{(i)}$	Initial Price $z^{(i)}$	Estimated Price $f^{(i)}$	ID	$x^{(i)}$	$y^{(i)}$	Initial Price $z^{(i)}$	Estimated Price $f^{(i)}$
1	1920	5083	22404	22681	26	9734	1680	24681	24896
2	5458	4662	11594	16445	27	5825	3803	15433	15012
3	7312	9171	22324	23571	28	4698	7768	29799	29876
4	4842	8802	31166	29958	29	2548	1917	27046	27479
5	7202	2609	19789	19887	30	2200	8481	28521	28043
6	3146	1038	30262	29180	31	3714	9810	31686	31370
7	4885	7378	30895	31596	32	8808	9167	24763	24689
8	6034	9311	18118	19224	33	3208	5536	23716	23453
9	4707	3497	27415	26452	34	1249	68	28471	24837
10	5368	4742	12639	16037	35	6303	4258	14267	14289
11	6068	8026	17541	18006	36	2531	541	29385	27826
12	1468	9462	27295	27160	37	1109	9604	28370	28565
13	821	8074	24943	25929	38	4165	9888	32979	32114
14	8001	1151	22075	22061	39	9343	2805	21007	20987
15	8491	9110	24101	24032	40	1572	1621	26527	26501
16	5615	8308	16985	18531	41	1892	33	28389	24689
17	8548	4353	19193	19383	42	6619	7038	15585	14808
18	9028	8511	21608	21892	43	4589	7755	30816	30521
19	6636	1911	17188	17023	44	2499	3446	24379	24524
20	7076	8627	20854	20439	45	1242	9032	28011	27971
21	6709	9056	21147	19871	46	6902	8752	18481	18942
22	6159	5741	13150	13313	47	6844	4143	15651	15923
23	9866	2409	22286	21712	48	7637	5303	16610	16570
24	904	8026	26195	26886	49	1504	5517	22432	22358
25	2890	856	29351	26927	50	4194	2871	27585	27082

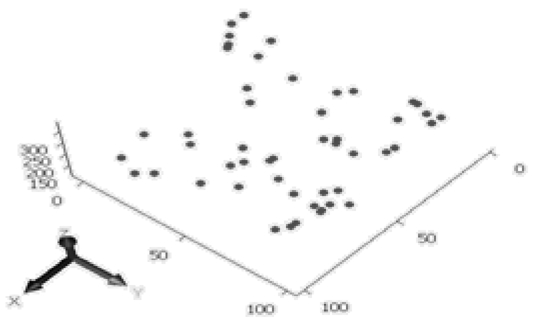

Fig. 5 The original cloud of 50 points

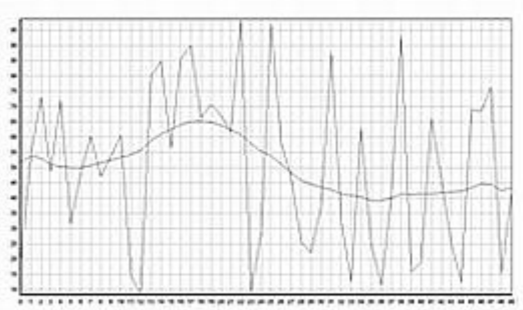

Fig. 6 Principal component on a grid line

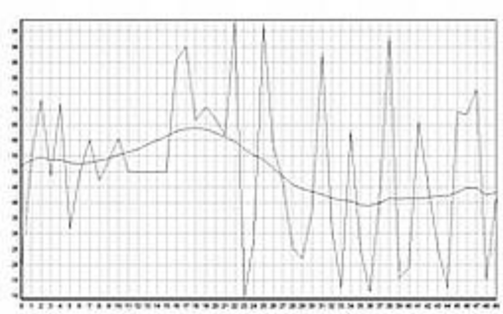

Fig. 7 The grid line of Fig. 6 with missing values

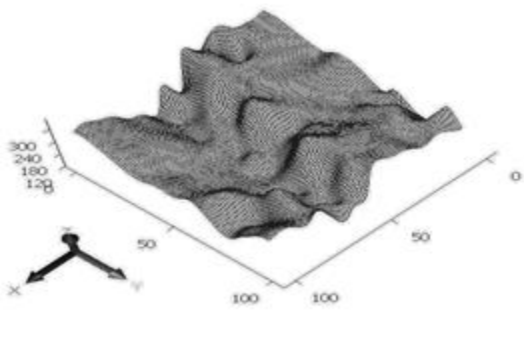

Fig. 8 Reconstructed interpolation grid after SSA

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 10, April 2014

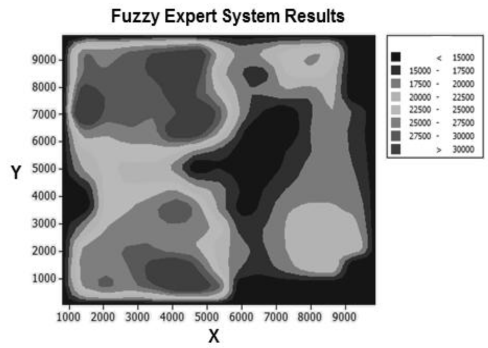

Fig. 8 Contour plot of the expert system answers about the land prices

V. CONCLUSION

In this paper, an approach to induce fuzzy rules from numerical data was presented. We have shown the problems and limitations of cluster-based and pattern-based fuzzy rule induction approaches. Our approach is based on multidimensional interpolation to induce fuzzy rules from a properly constructed approximation grid. While many interpolation methods are computationally efficient up to three dimensions, our approach based on iterative SSA procedures conducted on separate dimensions, can be applied to any dimensions without dramatically increasing the computations. This kind of fit is important in order to induce overlapping fuzzy rules and finally automatically obtain a self-learned rule-based inference system out of the available sample data. The weakness of clustering in case of missing data has been overcome and the method converges quickly. The use of fuzzy inference further improves estimations by accounting for the contribution of neighboring data without assuming any of it to be accurate and representative of the area it belongs. Moreover, by employing SSA upon the approximation grid before deriving fuzzy rules, our approach is applicable even when there is considerable noise and missing values in the data. Thus, the result of a rule based inference system, as demonstrated in the application example, is less prone to propagate errors in the data.

REFERENCES

[1] Zadeh, L. (2008). Is there a need for fuzzy logic? Information Sciences, 178(13), 2751-2779.

[2] Sato-Ilic, M., & Jain, L. (2006). Innovations in fuzzy clustering: theory and applications. Annals of Physics, Studies in Fuzziness and Soft Computing (Vol. 54, p. 258). Springer-Verlag New York Inc.

[3] Wang, L., & Mendel, J. (1992). Generating fuzzy rules by learning from examples. IEEE Transactions on systems, man and cybernetics, 22(6), 1414–1427.

[4] Casillas, J., Cordón, O., & Herrera, F. (2000). Improving the Wang and Mendel's fuzzy rule learning method by inducing cooperation among rules. In Proceedings of the 8th Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference (p. 1682–1688).

[5] Hong, T., & Lee, C. (1996). Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets and Systems, 84(1), 33-47.

[6] Iske, A. (2004). Multiresolution methods in scattered data modelling. Lecture Notes in Computational Science and Engineering. Springer Verlag.

[7] Schoellhamer, D. (2001). Singular spectrum analysis for time series with missing data. Geophysical Research Letters, 28(16), 3187–3190

[8] Ying, H. (1998). General SISO Takagi-Sugeno fuzzy systems with linear consequent are universal approximators, IEEE Trans. Fuzzy Syst., vol. 6, no. 4, pp. 582-587.

[9] Zeng K., Zhang N. Y. & Xu W. L. (2000). A comparative study on sufficient conditions for Takagi-Sugeno fuzzy systems as universal approximators, IEEE Trans. Fuzzy Syst., vol. 8, no. 6, pp. 773-780.

[10] Yang, C., Kao, S., Lee, F., & Hung, P. (2004). Twelve different interpolation methods: A case study of Surfer 8.0. In Proceedings of the XXth ISPRS Congress (Vol. 35, p. 778–785).

[11] Golyandina, N., Nekrutkin, V., & Zhigljavsky, A. (2001). Analysis of Time Series Structure: SSA and related techniques. Techno metrics, Monographs on Statistics and Applied Probability (Vol. 90). Chapman & Hall/CRC.

[12] Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modelling and control, IEEE Transactions on Systems, Man, and Cybernetics, vol. 15, no. 1, pp 116-132.

[13] Zeng, K., Zhang, N.Y. & Xu, W.L. (2000). A comparative study on sufficient conditions for Takagi-Sugeno fuzzy systems as universal approximates. IEEE Trans. Fuzzy Syst., vol. 8. no. 6, pp. 773-780.

[14] Tanaka, K., & Wang, H.O. (2001). Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach, New Work, John Wiley & Sons.

[15] Feng, G. (2010). Analysis and Synthesis of Fuzzy Control Systems: A Model-Based Approach}, CRC Press, Boca Raton, FL, 2010.