s c c h
software competence center
hagenberg

FUZZY.LOGIC.LAB.LINZ

# Advances in Knowledge-Based Technologies

Proceedings of the
Master and PhD Seminar
Winter term 2010/11, part 2

Softwarepark Hagenberg
SCCH, Room 0/2
15 February 2011

Software Competence Center Hagenberg
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 800
Fax +43 7236 3343 888
www.scch.at

Fuzzy Logic Laboratorium Linz
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 431
Fax +43 7236 3343 434
www.flll.jku.at

# Program

## Session 1. Chair: Bernhard Moser

13:00 Carlos Cernuda:
Takagi-Sugeno Fuzzy Inference Systems in Multivariate Calibration FLEXFIS

13:30 Edwin Lughofer:
On-Line Elimination of Local Redundancies in Evolving Fuzzy Systems

14:00 Henrike Stephani:
Hierarchical Clustering of High Volume Hyperspectral Terahertz Data Including Spacial and Spectral Features

## Session 2. Chair: Roland Richter

14:45 Jean-Luc Bouchot:
Automatic quality control for regularly textured surfaces based on template matching and discrepancy norm

15:15 Sándor Jenei:
An axiomatic treatment of equality

15:45 Thomas Vetterlein:
The versatility of the Logic of Approximate Entailment

# Takagi-Sugeno Fuzzy Systems in Multivariate Calibration FLEXFIS: FLEXible Fuzzy Inference Systems

# AdvKBT Winter 2010/11, 2

Carlos Cernuda[a]

[a]*Fuzzy Logic Lab Linz (FLLL)*
*Department of Knowledge-Based Mathematical Systems*
*Johannes Kepler University Linz, Austria*

## Abstract

Nowadays Multivariate Calibration has become a crucial factor in the industry, e.g. in non-destructive quality control. The advances in technology permit the obtaining, storing and handling of data sets with thousands of predictive variables, and, mainly due to economic and time restrictions, it is preferable to use as less instances as possible in the calibration process, leading to highly unbalanced data sets, as well as to have the possibility of employing evolving models. Classical methods, such as Principal Components Analysis(PCA) and Partial Least Squares(PLS), have been employed successfully during last decades. More recently, Ridge Regression or Support Vector Machines(SVM) have also been applied in Multivariate Calibration. But because of this new technical environment, there is a necessity of new techniques with the capability of covering existing gaps. Here is where Fuzzy Systems, and specifically FLEXFIS, have great chances.

*Keywords:* Multivariate Calibration, Takagi-Sugeno Fuzzy Systems, FLEXFIS, Forward Selection, Vector Quantization, incremental clustering.

## 1. Introduction: Motivation and State-of-the-Art

Chemometrics is the science of extracting information from chemical systems by data-driven means. It is applied to solve both descriptive and predictive problems in chemistry. In descriptive applications, properties of chemical systems are modeled

*Email address:* `carlos.cernuda@jku.at` (Carlos Cernuda) with the intent of learning the underlying relationships and structure of the system (i.e., model identification). In predictive applications, properties of chemical systems are modeled with the intent of predicting new properties or behavior of interest. We will focus on the latter.

Many chemical problems and applications of Chemometrics involve calibration. Calibration consists on connecting one (or more) sets of variables together. Usually one set (often called a *block*) is a series of physical measurements, such as some spectra or molecular descriptors, and the other contains one or more parameter such as the concentrations of a number of compounds or biological activity. The objective is develop models which can be used to predict properties of interest based on measured properties of the chemical system, such as pressure, flow, temperature, infrared, Raman, Near Infrared(NIR) spectra and mass spectra. Examples include the development of multivariate models relating 1) multi-wavelength spectral response to analyte concentration, 2) molecular descriptors to biological activity, 3) multivariate process conditions/states to final product attributes. The process requires a training data set, which includes reference values for the properties of interest for prediction, and the measured attributes believed to correspond to these properties. For case 1), for example, one can assemble data from a number of samples, including concentrations for an analyte of interest for each sample (the reference) and the corresponding infrared spectrum of that sample. Multivariate calibration techniques are then used to construct a mathematical model that relates the multivariate response (spectrum) to the concentration of the analyte of interest, and such a model can be used to efficiently predict the concentrations of new samples.

Due to the enormous number of predictive variables available in the actual data sets (e.g. tens of thousands in NIR data) it is desirable to reduce this number trying to keep as much information as possible. Thus dimensionality reduction and incremental learning approaches are needed as a complement to the inference process.

This paper is structured as follows: in section 2 we introduce Takagi-Sugeno(TS) Fuzzy Inference Systems(FIS), FLEXible Fuzzy Inference Systems(FLEXFIS) and Forward Selection(FS) algorithm for dimensionality reduction. Section 3 presents the description of the data set used in the experimentation, the State-of-Art methods used for comparison, and the validation scheme employed. In section 4 the results obtained are shown and commented. Finally, section 5 gives a general overview and sets the trend for future works.

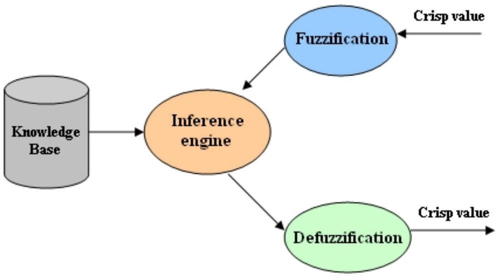


Figure 1: Modules of a FIS.

## 2. Methodology

### *2.1. Takagi-Sugeno Fuzzy Inference Systems*

A Fuzzy Inference System is a way of mapping an input space to an output space using fuzzy logic. A FIS tries to formalize the reasoning process of human language by means of fuzzy logic (that is, by building fuzzy IF-THEN rules). In general, a fuzzy inference system consists of four modules, depicted in Figure 1:

**Fuzzification Module:** Transforms input crisp values into membership degrees values.

**Knowledge Base:** Stores the IF-THEN rules, provided by experts and/or induced by data.

**Inference Engine:** Simulates human way of reasoning to infer a consequence regarding the inputs.

**Defuzzification Module:** Transforms the fuzzy consequence into a crisp value (not necessary in TS-FIS).

There are several types of IF-THEN rules that can be employed, Mamdani rules [10], Takagi-Sugeno rules [14], weighted rules, restricted rules, etc. We will focus on TS rules, that are rules where the consequence is a linear combination of the inputs. The advantage of TS-FISs is their capability of adaptation (flexibility) that facilitates an incremental learning process. A FIS is easily adaptable to changes in the modeled system just by adding new rules, or modifying or deleting existing ones. Moreover, they are able to approximate any nonlinear relationship to a certain degree of accuracy, thus being preferable to other fuzzy systems.

### *2.2. FLEXFIS*

FLEXFIS [9] is an approach for data-driven incremental learning for TS-FIS. The fuzzy sets used are Gaussian because they present several advantages (differentiability, full coverage, easy extraction of parameters to define them), even when they are more complex than triangular or trapezoidal fuzzy sets. Then, considering the input variables $X = (x_1, \ldots, x_p)$, a single output value $y$ and $C$ rules, our fuzzy model would be

$$\widetilde{y} = \widetilde{f}(X) = \sum_{i=1}^{C} l_i \Psi_i(X)$$

with the normalized membership functions

$$\Psi_i(X) = \frac{\exp\left[-\frac{1}{2}\sum_{j=1}^{p}(x_j - c_{ij})^2/\sigma_{ij}^2\right]}{\sum_{k=1}^{C}\exp\left[-\frac{1}{2}\sum_{j=1}^{p}(x_j - c_{kj})^2/\sigma_{kj}^2\right]}$$

and the consequent functions

$$l_i = w_{i0} + w_{i1}x_1 + \cdots + w_{ip}x_p$$

where $x_j$ denotes the $j$-th input variable, and $c_{ij}$ and $\sigma_{ij}$ denote respectively the center and the width of the Gaussian fuzzy set in the $j$-th premise part of the $i$-th rule.

During the learning process, three components are evolved in an incremental manner:

**Linear consequent parameters ($w_{i0}, w_{i1}, \ldots, w_{ip}$):** appearing in the rules consequents as output weights.

**Nonlinear premise parameters ($c_{ij}$ and $\sigma_{ij}$):** appearing in the input membership functions.

**Rule base:** concerning the number of rules, $C$, and the number of fuzzy sets per input dimension.

There are two possible approaches when it comes to learning the linear consequent parameters: global learning or local learning. The latter (in a recursive weighted least squares formulation ) is preferable to avoid numerical instability problems and also to get higher flexibility and a faster computational performance. The way to learn the premise parameters and to evolve the rule base is done by means of clustering on the input/output space, that is $(p + 1)$-dimensional. The clusters are projected

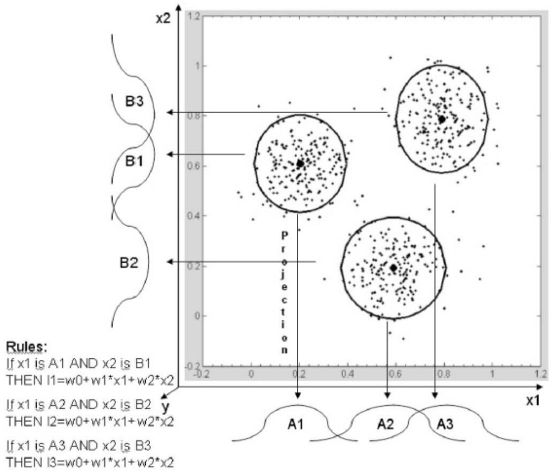


Figure 2: Projection of the three clusters onto the input axes $x_1$ and $x_2$.

onto the input axes, generating Gaussian fuzzy sets in the premise part of the rules. An example is depicted in Figure 2

Once we have new incoming samples, the update of cluster centers and widths, and/or the delation or creation of new ones is made based on a variant of Vector Quantization(VQ) [4] that includes a parameter, named *vigilance*, which controls the trade-off between generating new clusters or updating the existing ones, i.e., stability vs plasticity. The higher the vigilance, the harder to create a new center. In this way, we can take advantages of three characteristics:

- We can build up clusters in an incremental manner, without prefixing a number of clusters.

- We can synchronize the calculation of the ranges of influence of the clusters (approximating the variance of the data points belonging to the cluster) and the centers. Those ranges of influence will be used as widths for the Gaussian fuzzy sets projected.

- We have the possibility of calculating not only the distances to the centers of the clusters (like standard VQ does), but also to their surfaces, in order to look for the nearest (winning) cluster to the new incoming sample. If the point is inside an existing cluster, we calculate the distances to the center, and if it is outside all of them then we can decide if we prefer to calculate distances to the centers or to the surfaces. This last option would prevent from creating new clusters inside existing ones.

The center update is ruled by

$$\overrightarrow{c}_{win}^{(new)} = \overrightarrow{c}_{win}^{(old)} + \eta_{win}\left(X - \overrightarrow{c}_{win}^{(old)}\right)$$

where $\overrightarrow{c}_{win}$ is the center of the cluster nearest to the incoming data point $X$ in the $(p+1)$-dimensional input/output space, and $\eta_{win}$ is the learning gain for the mentioned cluster. The value of $\eta_{win}$ is crucial. A good value, based on empirical studies, is $\eta_{win} = \dfrac{1}{2k_{win}}$, with $k_{win}$ being the number of data points belonging to cluster *win*. The criteria for creating a new cluster is based on the vigilance, $\rho$. The point $X$ becomes the center of a new potential cluster if its distance to the winner cluster is at least $\rho$. The term *potential* indicates the necessity of waiting for some points to join that potential cluster before it is considered an actual cluster. To prevent from inappropriate $\rho$ values, that would lead to inappropriate clusters, a strategy based on merging close clusters and/or split big clusters and compare with the original situation in terms of a cluster partition quality measure is applied [8],[6]. The range of influence of a cluster is estimated by means of the variance of the data belonging to the cluster. In incremental mode, this estimation is given by the so-called recursive variance formula [11]

$$k_{win}\sigma_{win,j}^2(new) = (k_{win}-1)\,\sigma_{win,j}^2(old) + k_i\Delta c_{win,j}^2 + \left(c_{win,j}^2(new) - x_j\right)^2$$

where $\Delta c_{win,j}$ is the distance of the old prototype $c_{win,j}(old)$ to the new prototype $c_{win,j}(new)$ of the cluster nearest to the current point $X$ in the $j$-th dimension, and $k_{win}$ is the number of data points lying nearest to cluster $c_{win}$, which can therefore be simply updated through counting.

It is necessary to connect adequately the three components which are evolving, that is the premise parameters, the rule learning and the consequent parameters. We could update the premise part and then update the consequent part corresponding to the winner cluster, or viceversa. When we update the premise part two possibilities may arise:

- If a new cluster needs to be set, then a new rule may appear soon if some points reach the range of influence of this new potential cluster. The initial center is the incoming data point and the initial width must be small in order to avoid numerical instabilities. We update the center and width, and determine the parameters for the consequent, by means of recursive weighted least squares [7].

- If no new cluster needs to be set, the center and width of the winning cluster are updated and the parameter for the consequent are determined.

The reader should notice that recursive weighted least squares converges to the optimal set of parameters for the consequent part if the weights of the older points remain unchanged, but this is not the case for the rule belonging to the winning cluster for a new incoming data point as we are moving the fuzzy sets in the antecedent part. To palliate this problem a correction vector (for the linear parameter update) and a correction matrix (for the Hessian matrix involved in the quadratic optimization problem) are introduced in order to balance out the non-optimal situation toward the optimal one with respect to the degree of change in the antecedent part. Details about these corrections as well as the complete algorithms of both extended incremental VQ(Ext-Inc-VQ) and FLEXFIS are available in [9].

### *2.3. Dimensionality Reduction by Forward Selection*

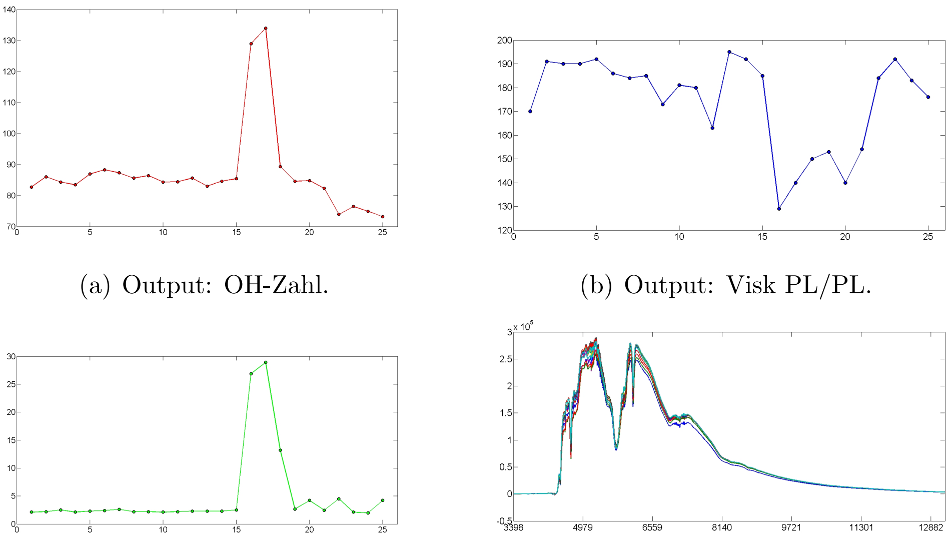
Forward Selection(FS) [12] is a simple data-driven model building approach. In this approach, one adds variables to the model one at a time. At each step, each variable that is not already in the model is tested for inclusion in it. The most significant of these variables, in terms of p-value, is added and its importance is stored. The contribution of the included variable is subtracted, and the next most significant variable is found. At the end all variables are ordered in terms of their significance, therefore we can select the $k$ most important ones, leading to a reduced data set with lower dimensionality which represents the original one.

Note that this multiple use of hypothesis testing means that the real type I error rate for a variable (i.e., the chance of including it in the model given it isn't really necessary), does not equal the critical level we could calculate over the original set. In fact, because of the complexity that arises from the complex nature of the procedure, it is essentially impossible to control error rates and this procedure must be viewed as exploratory, i.e., as a tool for ranking the variables.

## 3. Experimental Setup

### *3.1. Data set description*

The KCCS data set consists on 25 multiple instances, i.e. every instance is a set of 20 consecutive measurements, corresponding to 25 concrete moments. The type



(a) Output: OH-Zahl.

(b) Output: Visk PL/PL.

(c) Output: SZ.

(d) Inputs.

Figure 3: Graphic description of KCCS data set.

of data is NIR spectral data, considering 3101 equidistant wavelengths from 3398.46 to 13201.83. Therefore we are handling a matrix with 25 blocks of 20 rows and 3101 columns.

We consider three independent targets corresponding, in each of the moments, to

- The *hydroxyl number* (OH-Zahl).
- The *viscosity* (Visk PL/PL), obtained by means of a Parallel Plate Viscometry procedure.
- The *acidity* (SZ).

Figure 3 shows respectively the target values for the 25 multiple instances, in 3(a), 3(b) and 3(c), and the 25 mean values as a representation of the multiple instances, in 3(d).

Due to expert knowledge, the presence of outliers can be discarded, and the lowest and highest wavelengths can be omitted because they are only representative of noise. This can be detected just noticing how low the height is in both tails in figure 3(d). The key point is the decision of the cut levels on both tails. For that

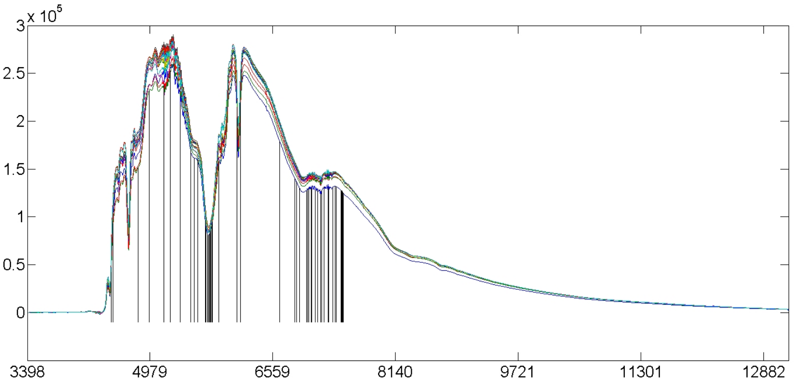


Figure 4: Most relevant variables according FS (black bars).

purpose, and adopting a conservative criterion, it is considered the ranking provided by FS algorithm with the objective of respecting the most important variables. Due to the difference in the slope in both sides, we opt for defining the cut points independently. Figure 4 shows the position of the most relevant predictor variables. The final decision is $\alpha_{left} = 40000$ and $\alpha_{right} = 128000$, in order to avoid discarding any important variable and not to take the risk of reducing too much (conservative criterion). The reduction in the number of wavelengths is from 3101 to 982.

### *3.2. State-of-art methods used for comparison*

It is known that Multiple Linear Regression (MLR) suffers from two different problems

- The abundance of response variables relative to the number of available samples which leads to an undetermined situation. This is the case in KCCS data set.
- The possibility of collinearity of the response variables, which leads to unstable matrix inversions and unstable regression results.

These problems can be dealt by means of other kinds of regression, like Principal Components Regression(PCR), Partial Least Squares(PLS) or Ridge Regression introducing a penalty term, see [13, 5, 12]. Principal Components Analysis(PCA) finds combination of variables that describe major trends in the data. If $X$ is a data matrix with $m$ rows and $n$ columns, and with each variable being a column and each sample a row, PCA decomposes $X$ as the sum of $r$ $t_i$ - $p_i$ inner products, where $r$ is the rank of the matrix $X$:

$$X = t_1 p_1^T + t_2 p_2^T + \ldots + t_k p_k^T + \ldots + t_r p_r^T$$

where $r$ must be less than or equal to the smaller dimension of X, i.e. $r = \min\{m, n\}$. The $t_i$ vectors are known as *scores* and contain information on how the samples relate to each other. The $p_i$ vectors are known as *loadings* and contain information on how the variables relate to each other. The $t_i$ - $p_i$ pairs are ordered by the amount of variance captured. The sum is considered up to $k$ components, being the rest consolidated into a residual matrix. The determination of $k$ is crucial in order to capture enough information. The usual way consists on calculate the eigenvalues, which indicate the amount of information captured.

In PCR, instead of regressing the properties of interest onto a set of the original response variables, the properties are regressed onto the $k$ principal component scores of the measured variables,which, by definition, are orthogonal, and therefore well-conditioned.

Partial Least Squares(PLS) regression is related to both PCR and MLR, and can be thought of as stated in between them. The former finds factors that capture the greatest amount of variance in the predictor variables while the latter seeks to find a single factor that best correlates predictor variables with predicted variables. PLS attempts to find factors which both capture variance and achieve correlation by means of projecting not only the predictor variables (like PCA), but also the predicted ones, to new spaces so that the relationship between successive pairs of scores is as strong as possible.

Considering matrix notation, we would have

$$X \;=\; TP \;+\; E \quad , \quad y \;=\; Tq \;+\; f$$

where $T$ is the scores matrix, $P$ and $q$ are the loadings matrix and loadings vector for the inputs $X$ and the output $y$ respectively, and $E$ and $f$ represent the residual errors in both projections. In summary, PLS consists on a search of directions in the factor space that are associated with high variation in the responses but biasing them toward directions that are accurately predicted.

Another choice is the use of Support Vector Regression(SVR). It is a nonlinear method and its use will be good in order to compare with FLEXFIS, which is

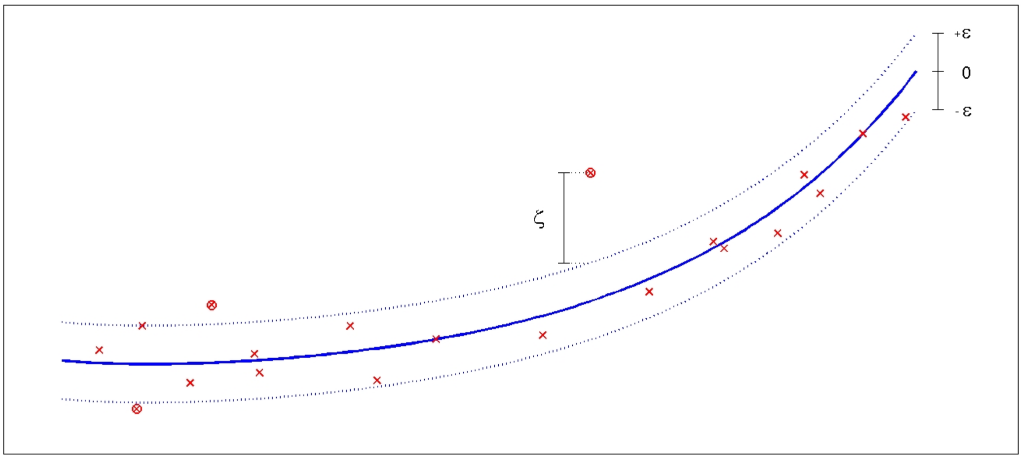


Figure 5: Only circled points obtain non-zero loss value.

also nonlinear. We consider two well known techniques: $\epsilon$-SVR and $\nu$-SVR. The basic idea consists on stating a tolerance $\epsilon$ and then looking for a function $f(x)$ that has at most $\epsilon$ deviation from the actually obtained targets for all the training data, and at the same time is as flat as possible, i.e., we do not care about errors as long as they are less than $\epsilon$ but we do not accept any deviation larger than this [1]. More in detail, let the training set be $D = (X|y)$. We denote the i-th row of $X$ by $X_i$ and the i-th output value by $y_i$. Therefore, the i-th row of $D$ is $(X_i, y_i)$, for $i = 1, \dots, n$. In $\epsilon$-SVR, $X$ is firstly mapped to $Z = \phi(X)$ in feature space $\mathfrak{F}$ and then a linear function

$$f(X, W) = W^T Z + b$$

is constructed in $\mathfrak{F}$ so that its deviation from the training data is minimum, according to the $\epsilon$-insensitive loss function $\zeta$ [2]

$$\zeta(X) = |y - f(X)|_\epsilon = \begin{cases} 0 & \text{if } |y - f(X)| \leq \epsilon \\ |y - f(X)| - \epsilon & \text{otherwise} \end{cases}$$

that considers zero any error below $\epsilon$, while at the same time is as flat as possible, as is depicted graphically, in a toy example, in Figure 5.

This idea can be formulated in terms of a programming problem [15]

$$
\begin{aligned}
&\text{minimize} \\
&\qquad \tfrac{1}{2}||W||^2 + C\textstyle\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\
&\text{subject to} \\
&\qquad \left\{ \begin{array}{rcl} y_i - f_i & \leq & \epsilon + \xi_i^* \\ f_i - y_i & \leq & \epsilon + \xi_i \\ \xi_i, \xi_i^* & \geq & 0 \end{array} \right. \qquad 1 \leq i \leq n
\end{aligned}
$$

where $C > 0$ is a user defined constant which determines the trade-off between the flatness of $f$ and the amount up to which deviations larger than $\epsilon$ are tolerated. The value of $\epsilon$ must be prefixed. $\nu$-SVM allows us to determine it automatically using an additional constant $\nu \geq 0$ to trade-off $\epsilon$ against training accuracy and model complexity.

In terms of programming

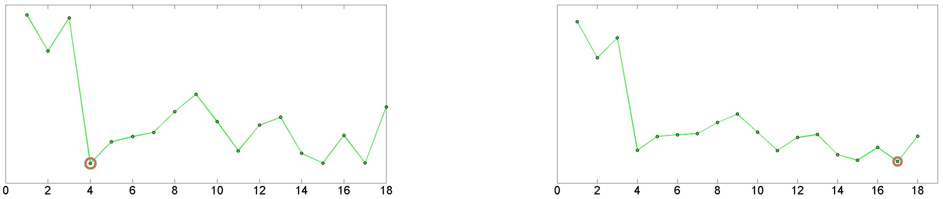
$$
\begin{aligned}
&\text{minimize} \\
&\qquad \tfrac{1}{2}||W||^2 + C\left(\nu\epsilon + \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right) \\
&\text{subject to} \\
&\qquad \left\{ \begin{array}{rcl} y_i - f_i & \leq & \epsilon + \xi_i^* \\ f_i - y_i & \leq & \epsilon + \xi_i \\ \xi_i, \xi_i^* & \geq & 0 \end{array} \right. \qquad 1 \leq i \leq n
\end{aligned}
$$

These problems can be transformed into quadratic programming problems utilizing Lagrange multipliers and dual programming [3]. The result is a formulation in terms of scalar products that can be extended to a non-linear case by means of using different kernel functions and applying the so called *kernel trick*. In our case, we opt for the widely used Radial Basis Function(RBF) kernel.

### *3.3. Validation scheme*

Taking into account that we have repeated measures, we treat all those 20 repeated measures of every multiple instance as one. Therefore, all 20 measures corresponding to a multiple instance will never be split into a training and testing set, to avoid a too optimistic prediction. Due to the reduced number of available instances (only 25), it is not possible to take apart a validation set and use the rest for modeling.

The only sensible option is to perform a cross validation(CV) over the whole



(a) Penalized RMSE.

(b) Non-penalized RMSE.

Figure 6: Example of the behavior of $RMSE_{\alpha}^{pen}$, on FLEXFIS+FS for SZ.

data set. We opt for a leave-3-out CV to leave out 12% of the instances. The CV is made 10 times, calculating the average root mean squared error(RMSE) and the standard deviation(SD) obtained in the 10 shuffles for every algorithm.

The way to determine the optimal number $a$ of principal components in PCR, the optimal number $a$ of latent variables in PLS regression and the optimal dimensionality in FLEXFIS will be also based on the RMSE of CV, but penalizing high complexity. This penalized error is given by

$$RMSE_{\alpha}^{(pen)} = RMSE(1 + \alpha\ dim)$$

Notice that this penalized error prioritizes RMSE over dimensionality. Empirically, we have determined $\alpha = 0.04$ as an adequate value. As an example, figure 6 shows the difference in the optimal dimensionality with, 6(a), or without, 6(b), penalization.

In $\epsilon$-SVR and $\nu$-SVR we prefix $\epsilon = 0.1$ and $\nu = 0.5$, that are standard values in the literature. Thus, there are two parameters to be tuned: the *cost* $C$ and the $\gamma$ parameter of RBF kernel. In order to determine their final values, a grid search(GS) is done, using the ranges for the parameters $C \in \{2^{-5}, 2^{-3}, \ldots, 2^{15}\}$ and $\gamma \in \{2^{-15}, 2^{-13}, \ldots, 2^{3}\}$.

When it comes to FLEXFIS, we try to combine not only with FS, but also with PCA and PLS, in order to compare their performances when combined with FLEXFIS. In all cases (FLEXFIS+FS, FLEXFIS+PCA and FLEXFIS+PLS) we need to tune 2 parameters, *vigilance* $v$ and *dimensionality* $d$. Again the chosen way is GS, considering $v \in \{0.1, 0.2, \ldots, 0.9\}$ and $d \in \{1, 2, \ldots, 18\}$.

## 4. Results

The targets are treated separately, even when OH-Zahl and SZ are softly related. The differences in the results obtained support this way of proceeding. Therefore we

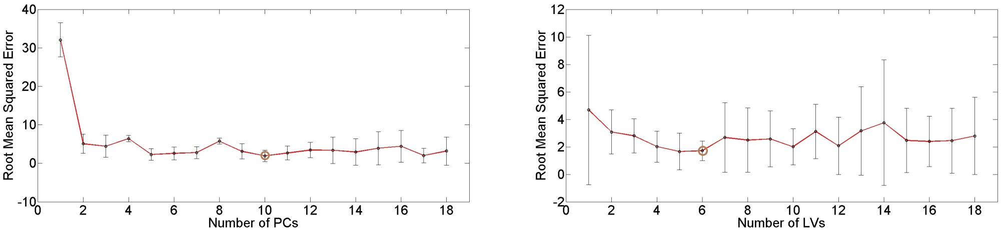


Figure 7: Selection of number of PCs (left) and LVs (right) for OH-Zahl (best marked).

present the results target by target.

### *4.1. OH-Zahl*

Figure 7 shows the CVRMSE corresponding to different numbers of principal components(PCs) and latent variables(LVs), from 1 to 18. The black bars indicate the standard deviation. The chosen values are $a = 10$ and $a = 6$ respectively.

In both $\epsilon$-SVR and $\nu$-SVR, there is no dimensionality reduction involved. Hence, the best values for the parameters are presented in the summary tables.

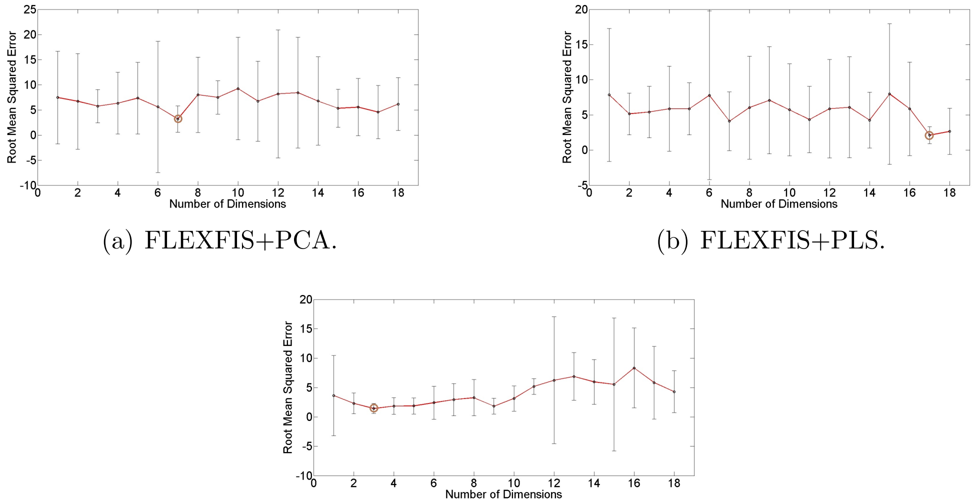
For FLEXFIS[1] algorithm, the errors obtained are depicted, for all three dimensionality reduction algorithms, in figure 8. The best combination of parameters for PCA is $v = 0.9$ and $d = 7$, for PLS $v = 0.4$ and $d = 7$, and $v = 0.9$ and $d = 3$ for FS.

Figure 9 presents an overview of the behavior of the algorithms which have to be with dimensionality reduction.

The results obtained, as well as the corresponding best parameters, are summarized in table 4.1.

First we should notice that SVR performance is much worse than the performance of any other method. The best algorithm is FLEXFIS+FS with PLS obtaining similar results, but with double dimensionality. Finally we can appreciate that both PCA and PLS behave better on their own than when combined with FLEXFIS. Taking into account the high unbalance between the number of predictor variables and the number of available instances, it is remarkable the low dimensionality obtained by FLEXFIS+FS wrt PLS.

[1] The reader should take into account that the graphics corresponding to FLEXFIS in combination with any algorithm show the CVRMSE and the standard deviation for the best vigilance in each dimensionality.



(a) FLEXFIS+PCA.

(b) FLEXFIS+PLS.

(c) FLEXFIS+FS.

Figure 8: FLEXFIS errors for OH-Zahl (best marked).

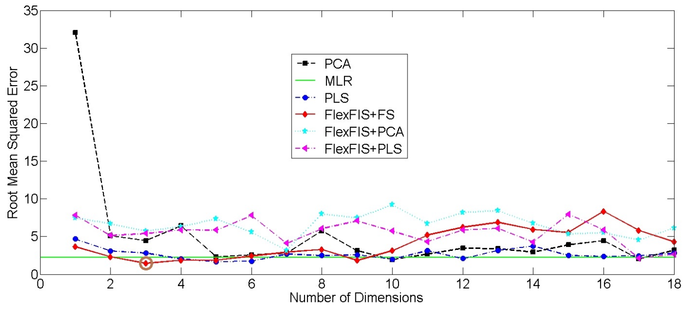


Figure 9: Comparison, for different dimensionality, over OH-Zahl (best marked).

| Algorithm | Parameters | RMSE | SD |
|---|---|---|---|
| **MLR** | — | 2.2546 | 1.4684 |
| **PCR** | $a = 10$ | 1.9346 | 1.4736 |
| **PLS** | $a = 6$ | 1.7183 | 0.7143 |
| **$\epsilon$-SVR** | $\epsilon = 0.1$, $C = 2^{-1}$, $\gamma = 2^1$ | 5.6991 | 10.6569 |
| **$\nu$-SVR** | $\nu = 0.5$, $C = 2^{-5}$, $\gamma = 2^3$ | 5.6502 | 10.4902 |
| **FLEXFIS+PCA** | $v = 0.9$, $d = 7$ | 3.1730 | 2.6116 |
| **FLEXFIS+PLS** | $v = 0.7$, $d = 17$ | 2.1065 | 1.2061 |
| **FLEXFIS+FS** | $v = 0.9$, $d = 3$ | **1.4294** | **0.8478** |

Table 1: Results for OH-Zahl (range = 60.74).

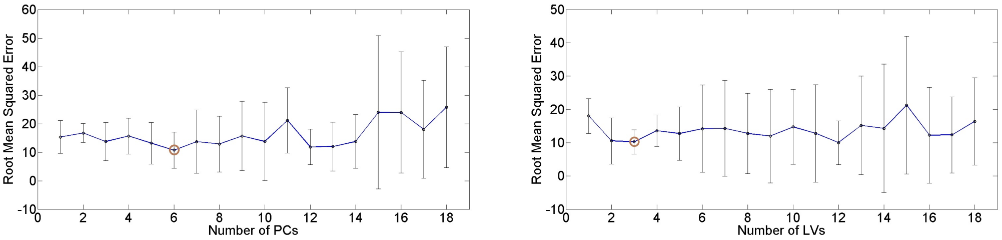


Figure 10: Selection of number of PCs (left) and LVs (right) for Visk PL/PL (best marked).

### *4.2. Visk PL/PL*

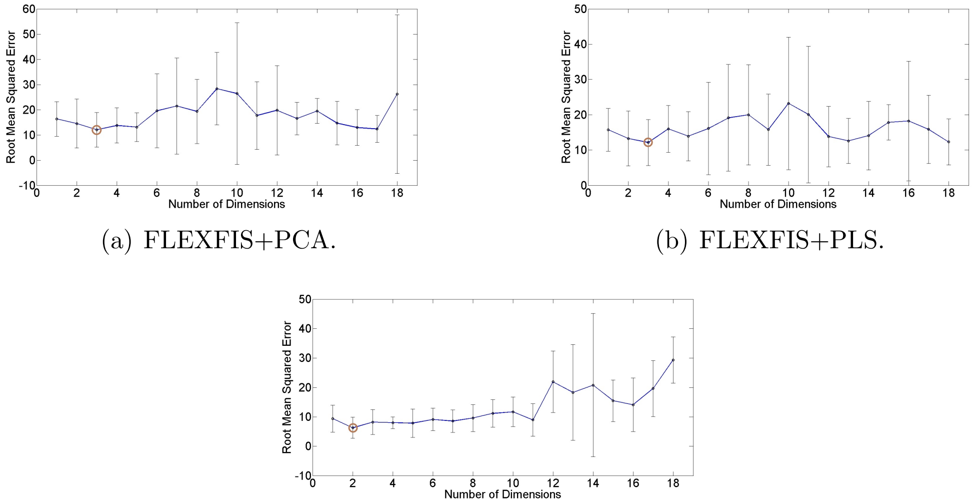
Figure 10 shows the CVRMSE corresponding to different numbers of principal components (PCs) and latent variables(LVs), from 1 to 18. The chosen numbers are $a = 6$ and $a = 3$ respectively.

When it comes to FLEXFIS, the errors obtained are depicted in figure 11. The best combinations of parameters are $v = 0.8$ and $d = 3$, $v = 0.9$ and $d = 3$, and $v = 0.8$ and $d = 2$ for PCA, PLS and FS respectively.

The overview of the behavior of the algorithms which have to be with dimensionality reduction is shown in figure 12.

Table 4.2 summarizes the results obtained as well as the corresponding best parameters.

Again SVR behavior is not good but the difference with the rest is not as big as before. The performance of all algorithms but FLEXFIS+FS is similar, both in RMSE and in dimensionality. FLEXFIS+FS is clearly better than the rest in both aspects.



(a) FLEXFIS+PCA.

(b) FLEXFIS+PLS.

(c) FLEXFIS+FS.

Figure 11: FLEXFIS errors for Visk PL/PL (best marked).

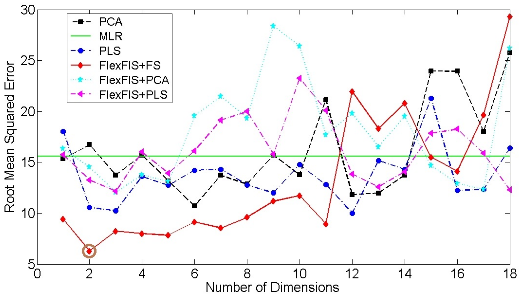


Figure 12: Comparison, for different dimensionality, over Visk PL/PL (best marked).

| **Algorithm** | **Parameters** | **RMSE** | **SD** |
|---|---|---|---|
| **MLR** | — | 15.5784 | 11.9403 |
| **PCR** | $a = 6$ | 10.1510 | 7.3332 |
| **PLS** | $a = 3$ | 10.2541 | 3.6207 |
| **$\epsilon$-SVR** | $\epsilon = 0.1$, $C = 2^{-3}$, $\gamma = 2^{-7}$ | 13.6281 | 11.3215 |
| **$\nu$-SVR** | $\nu = 0.5$, $C = 2^{5}$, $\gamma = 2^{-7}$ | 14.6692 | 6.8904 |
| **FLEXFIS+PCA** | $v = 0.8$, $d = 3$ | 12.0464 | 6.8625 |
| **FLEXFIS+PLS** | $v = 0.9$, $d = 3$ | 11.1374 | 6.8777 |
| **FLEXFIS+FS** | $v = 0.8$, $d = 2$ | **6.2785** | **3.5688** |

Table 2: Results for Visk PL/PL (range = 66).

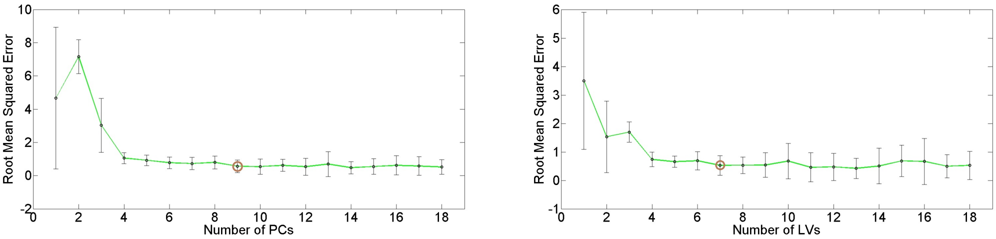


Figure 13: Selection of number of PCs (left) and LVs (right) for SZ (best marked).

### *4.3. SZ*

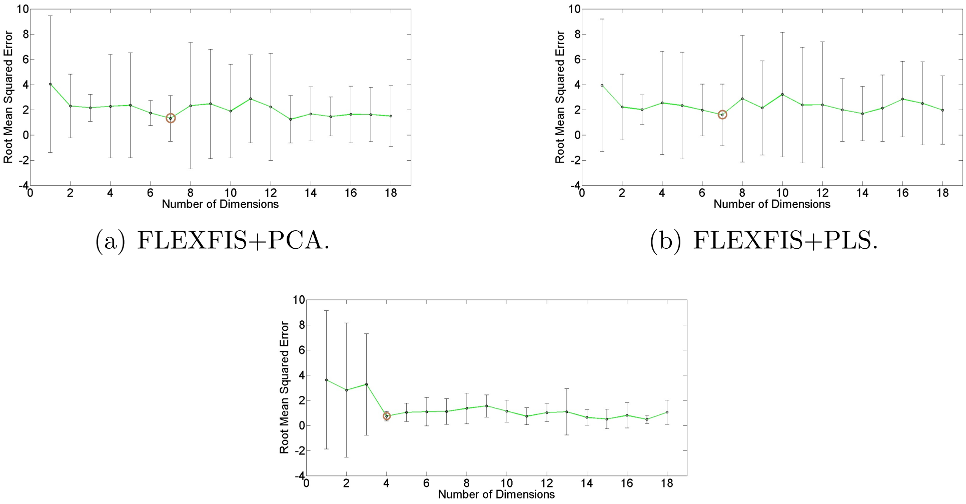
The CVRMSE corresponding to different numbers of principal components(PCs) and latent variables (LVs),from 1 to 18, can be seen in figure 13. The chosen numbers are $a = 9$ and $a = 7$ respectively.

The best combinations in the case of FLEXFIS are $v = 0.8$ and $d = 7$, $v = 0.8$ and $d = 7$, and $v = 0.9$ and $d = 4$, for PCA, PLS and FS respectively, as shown in figure 14.

Table 4.3 summarizes the results obtained as well as the corresponding best parameters.

The overview of the behavior of the algorithms which have to be with dimensionality reduction is shown in figure 15.

The bad performance of SVR persists, and this time also FLEXFIS+PCA and FLEXFIS+PLS show much worse results than PLS, PCR and FLEXFIS+FS. Those three obtain similar results, being PCR the best algorithm. Notice that the optimal dimensionality for FLEXFIS+FS is again the lowest.



(a) FLEXFIS+PCA.

(b) FLEXFIS+PLS.

(c) FLEXFIS+FS.

Figure 14: FLEXFIS errors for SZ (best marked).

| **Algorithm** | **Parameters** | **RMSE** | **SD** |
|---|---|---|---|
| **MLR** | — | 0.6936 | 0.4496 |
| **PCR** | $a = 10$ | **0.4722** | **0.3745** |
| **PLS** | $a = 13$ | 0.5241 | 0.3393 |
| **$\epsilon$-SVR** | $\epsilon = 0.1$, $C = 2^{-5}$, $\gamma = 2^{-11}$ | 3.1739 | 5.5629 |
| **$\nu$-SVR** | $\nu = 0.5$, $C = 2^{-5}$, $\gamma = 2^{-5}$ | 2.7015 | 5.8859 |
| **FLEXFIS+PCA** | $v = 0.8$, $d = 7$ | 1.3155 | 1.8110 |
| **FLEXFIS+PLS** | $v = 0.8$, $d = 7$ | 1.5978 | 2.4417 |
| **FLEXFIS+FS** | $v = 0.9$, $d = 4$ | 0.5363 | 0.3785 |

Table 3: Results for SZ (range = 26.92).

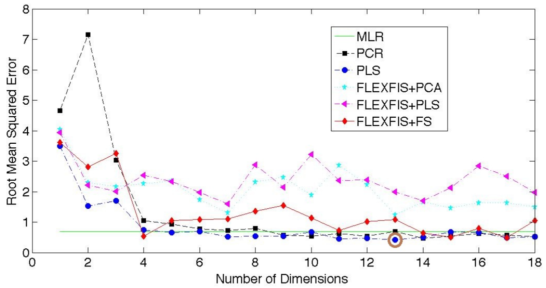


Figure 15: Comparison, for different dimensionality, over SZ (best marked).

## 5. Conclusion

We have presented Takagi-Sugeno Fuzzy Inference Systems in general, and FLEXFIS in particular, as a valid tool that can coexist with, and eventually overcome, the classical algorithms widely used in Multivariate Calibration. The possibilities are enormous due to the capability of evolution in a simple way, what would permit online mode MC, facilitated by the low dimensionality required. The applicability to real world problems has been proved in the experiments that have been carried out, where the unbalance between the number of predictor variables and the number of available instances is huge, as usual nowadays in industrial environments.

Another aspect that has been obviated in this paper is the inherent interpretability that a FIS, based in rules, has. This fact can make a non-expert final user feel more comfortable while using a FIS (grey box) instead of other opaque methods (black box).

## References

[1] B. Schölkopf A.J. Smola. A tutorial on support vector regression. *Neural and Computational Learning II*, (030), 1998.

[2] V. Vapnik C. Cortes. Support vector networks. *Machine Learning*, 20:273 – 297, 1995.

[3] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, New York, 1989.

[4] R. Gray. Vector quantization. *IEEE ASSP*, 1(2):4 – 29, 1984.

[5] T. Næs H. Martens. *Multivariate Calibration.* John Wiley & Sons, New York, 1989.

[6] E. Hllermeier J. Beringer. Online clustering of parallel data streams. *Data and Knowledge Engeneering*, 58(2):180–204, 2006.

[7] L. Ljung. *System Identification: Theory for the User.* Prentice-Hall, NJ, 1999.

[8] E. Lughofer. Extensions of vector quantization for incremental clustering. *Pattern Recognition*, 41(3):995–1011, 2008.

[9] E. Lughofer. FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models. *IEEE Trans. on Fuzzy Systems*, 16(6):1393–1410, 2008.

[10] E.H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Transactions on Computers*, 26(12):1182 – 1191, 1977.

[11] H. Yue S.Qin, W. Li. Recursive pca for adaptive process monitoring. *J. Process Control*, 10:471–486, 2000.

[12] J. Friedman T. Hastie, R. Tibshirani. *The Elements of Statistical Learning.* Springer, New York, 2001.

[13] H. Martens T. Næs. Principal component regression in nir analysis: Viewpoints, background details and selection of components. *Journal of Chemometrics*, 2, 1988.

[14] M. Sugeno T. Takagi. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics. Part C: applications and reviews*, 15:116 – 132, 1985.

[15] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer, New York, 1995.

# On-Line Elimination of Local Redundancies in Evolving Fuzzy Systems

Edwin Lughofer

## Abstract

In this paper, we examine approaches for reducing the complexity of fuzzy systems incrementally learned and evolved automatically from data streams (evolving fuzzy systems = EFS) by eliminating local redundancies. As data streams are typically recorded with a certain frequency and coming in in a sample-wise on-line fashion, the complexity reduction steps should support fast incremental single-pass processing steps (i.e. not requiring any re-building steps with all or a great portion of samples seen so far). In evolving fuzzy systems, such (permanent) reduction steps are important due to several reasons: 1.) originally distinct rules (representing distinct local regions in the input/output data space) may move together over time and get significantly over-lapping as data samples are filling up the gaps in-between these, 2.) two or several fuzzy sets in the fuzzy partitions may become redundant because of projecting high-dimensional clusters onto the single axes, 3.) they can be also seen as a first step towards a better readability and interpretability of fuzzy systems (as unnecessary information is discarded and the models are made more transparent). Two techniques for tackling the problem of eliminating redundancies in EFS for regression problems are presented. One is performing a new rule merging approach directly in the product cluster space using a novel concept for calculating the similarity degree between an updated rule and the remaining ones. The second one is operating directly in the fuzzy partition space, where redundant fuzzy sets are merged based on their joint $\alpha$-cut levels. Redundancy on feature as well as fuzzy partition level is recognized by a novel fuzzy inclusion measure. Merging or pruning of rules is decided based on the level of contradictions, measured in terms of dissimilarity of rule consequents compared to the similarity of their antecedent parts. The complexity reduction approaches are evaluated based on high-dimensional (noisy) real-world measurements, where *FLEXFIS* method serves as incremental learning engine for evolving the models. Based on this empirical comparison, it will be shown that the novel techniques are 1.) fast enough in order to cope with on-line demands and 2.) produce fuzzy systems with less structural components while at the same time still achieving similar accuracies as when not applying any reduction steps in EFS.

## Index Terms

evolving fuzzy systems, redundancy elimination, similarity, complexity reduction, on-line rule and fuzzy set merging

# Hierarchical Clustering of High Volume Hyperspectral Terahertz Data – Including Spacial and Spectral Features

Henrike Stephani

henrike.stephani@itwm.fraunhofer.de

## Abstract

Hyperspectral Terahertz time-domain spectroscopic images (THz-TDS images) usually have a high dimensionality i.e. consist in images with hundreds of channels in each pixel. We propose to reduce this dimensionality down to $\sim 20$ features that include a information about most of the compounds that can be detected and identified with the THz-TDS technology. Among these features are time-domain features as well as a set of wavelet-coefficients that represent the information contained in the amplitude spectrum.

In addition to the high dimensionality, hyperspectral imaging data is of high volume. Even with low resolution such, e.g. 200 x 100 pixel, the amount of spectra accumulates to 20 000. We therefore use preclustering and apply two different kinds of hierarchical clustering techniques on the remaining groups. First we use each groups mean sample and then apply standard Euclidean distance hierarchical average link clustering. In the second approach we apply the Chameleon algorithm.

To include spacial characteristics as well, we calculate a spacial similarity between the clusters in each step and include that into the clustering algorithm.

Technical Report
FLLL–TR–1101

# Automatic quality control for regularly textured surfaces based on template matching and discrepancy norm

15/02/2011

Jean-Luc Bouchot
Fuzzy Logic Laboratorium Linz
email: jean-luc.bouchot@jku.at
Gernot Stübl, Bernhard Moser
Software Competence Center Hagenberg
email {gernot.stuebl, bernhard.moser}@scch.at

**Abstract —** In this talk we are interested in automatic quality control of regularly textured surfaces. Besides describing the state of the art methods based on statistics and texture analysis, with their pros and cons, we will detail our new approach which is mainly based on a template matching algorithm.

The development of this algorithm makes a full use of the discrepancy norm being monotonic with respect to shifting. The principle of our novel method is to register an input patch into a bigger reference defect-free pattern. If this registration fails then we are in presence of a defect area. Moreover in order to improve the computational cost as well as the localization of the defect area, a coarse-to-fine approach is implemented.

**Key words —** *Discrepancy norm, automatic quality control, template matching, image registration*

Institut für Wissensbasierte Mathematische Systeme

Johannes Kepler Universität Linz
Altenberger Str. 69
A-4040 Linz
Austria

Fuzzy Logic Laboratorium Linz

Softwarepark 21
A-4232 Hagenberg
Austria

# 1 Introduction

Automatic defect detection is an important area in the quality control domain. It is the problem of finding defects an automatic way in production coming out on a conveyor. A human-based system is all but reliable, mainly for two reasons. First one can not guarantee a supervisor to be concentrated the whole day looking at a running conveyor and visual abilities decrease with the number of hours worked. Second, there are some defects that can not be seen by the human eyes and need some special optical set-ups to appear. Therefore a reliable and efficient algorithm to detect these defects is needed.

Many methods have been developed in past years as described in the recent surveys[Xie08, Kum08], but to our knowledge almost none of them are based on a template matching approach. It only has been studied in the context of rigid registration of ceramic tiles[CP00, LGALVGP01] and for inspection of wafers[XG00]. Template matching approaches seem to have some weakness when applied to high frequency images, as explained in our previous work[BSM11]. However, we base our approach on a registration based algorithm, making use of the monotonicity of the discrepancy norm.

This article is articulated as follows. First we give a short review of the state of the art in Section 2. Afterwards details on the studied approach are given in Section 3. This approach has been described our work[BSM11] and we are interested here in giving some results by comparing with other methods in Section 4. Finally Section 5 concludes this paper and gives an overview of what should be done next.

# 2 State of the Art

Many researches have been done in the past year on the automatic inspection fabric. We are in this article particularly interested in automatically finding defects in regularly textured surfaces. Different methods with different philosophies have been studied. We do not claim the following study to be exhaustive but we think the presented methods are rather accepted and reliable within the community.

First we would like to mention the statistical methods. These approaches rely on some statistical description of the analyzed texture. Firstly the first order statistics have been studied with histogram based descriptions[Ng07]. Then the Gray Level Co-occurrence Matrix appeared. It allows to describe the dependencies between two pixel intensities' values. The decision whether an area is defective or defect-free is done by computing some features out of this matrix. Among others, we can cite the entropy, local homogeneity and energy, which seem to be the one yielding the best results. We refer the reader to [HSD73] for more details.

Another aspect studied within the idea of detecting defects in textures is based on spectral representations of signals. The Fourier transform arrived first[SSG96], as it allows one to detect irregularity in the spectrum of the texture. However it did not allow the possibility to localize and detect the defect at the same time. Therefore methods based on Short Time Fourier Transform and Gabor filter[BBL02] have been developed. Since the 1990s most of the effort have been place on the study of wavelet based methods. Wavelets are well suited for this task as it allows to describe an image on a scale-time space yielding localization and the presence of a defect at the same time. We refer the reader to [LB97, NPSN05, MMMJ09] for more details.

Besides the example given in the introduction it seems that template matching methods are of no interests to the community working on automatic inspection. We can though think of an algorithm which would try to find matching pairs of repetitive patterns. Whenever a pattern is not known it would be classified as defective.

Before we describe our novel template based approach in the next section, we would like to add another remark. All the methods mentioned above are based on the fact that one can and will learn a classifier. In it clear that this classifier can also be a complete problem to solve, as one could choose a basic classifier such as a nearest neighbor, or think of more complex ones such as neural networks or support vector machines. We will anyway not give more details on this topic in this article.

# 3 Our approach

The algorithm we will describe here has been introduced first in[BSM11]. However, to keep this article self contained, we want to give the major details of the approach.

## 3.1 Idea

As suggested in the introduction (Section 1), the algorithm is based on a template matching approach. Assume we have a flow of images $I_k$ where $k$ represents the frame number. We are interested in telling whether a defect is present in the current frame or not. Assume moreover that we are given a reference defect-free pattern $R$. Our idea is to decompose our frame $I_k$ into small patches and to analyze each of them separately. This analysis process corresponds to some registration, as shown in Figure 1. Let $i_k(x, y)$ denote an $n \times m$ patch of image $I_k$ centered at $(x, y)$. For all values of $x, y \in Z$ we produce a dissimilarity value $\alpha$:

$$\alpha_k(x, y) = \min_{H \in \mathcal{H}} \|T(i_k(x, y)) - \widetilde{R}\|_D \tag{1}$$

where $\mathcal{H}$ denotes a set of admissible transformations, $\widetilde{R}$ is the restriction of $R$ to the support of $T(i_k(x, y))$ and $\| \cdot \|_D$ represents the discrepancy norm[Mos09, NW87].

This general description allows any kind of transformations to be considered (for instance blurring/deblurring to face vibrations) but we will consider only shifting which gives the following formulation for Equation 1:

$$\alpha_k(x, y) = \min_{t_x, t_y \in R} \|i_k(x, y) - R(x_0 - t_x, y_0 - t_y)\|_D \tag{2}$$

with $(x_0, y_0)$ being a reference point in the reference image $R$.

The set of all these values gives us an input map which can be interpreted as a *Local Dissimilarity Map (LDMap)* similar to the one described in[BNMR08]. Finally the dissimilarity values obtained are used to classify any single pixel into defective or defect free.

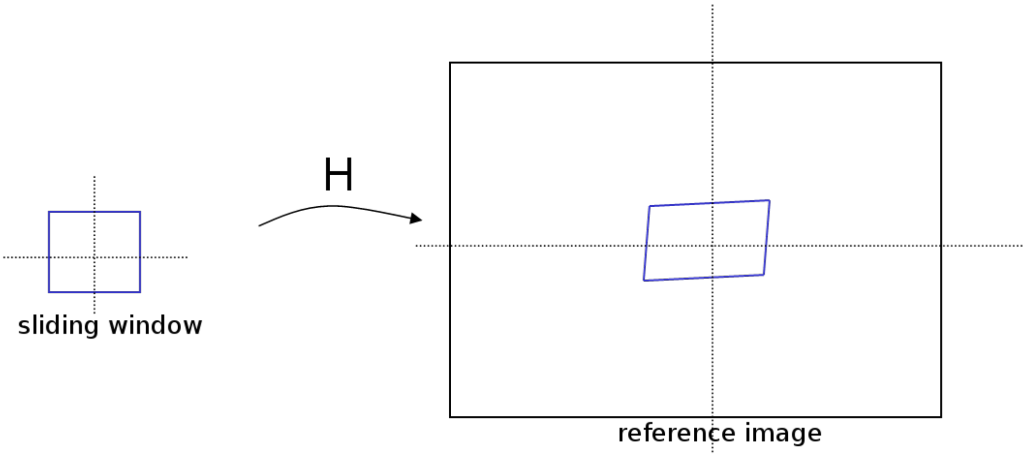


Figure 1: Description of the registration process. An input sliding window is fitted through the transformation H on the reference image

### 3.2 Improvements

In the algorithm described above, the resulting values in the LDMap are strongly correlated. Indeed whenever a complete patch can be without doubt (ie. with small dissimilarity value) registered on the reference, the probability that one of the pixels should be classified as a defective part of the surface is particularly small. The first improvement makes full use of this remark by implementing a coarse-to-fine approach. By doing this, we are able to classify rather big patches as defect-free patches efficiently and just where there are doubts, we will refine the localization.

Another idea is to make use of the particular norm we have chosen. It has been proven[Mos09] that for positive signals $I \geq 0$ the autocorrelation function based on the discrepancy norm is monotonic. We can write this in the following property: *Let* $\Delta_t(I) = \|I(\cdot - t) - I\|_D$ *be the autocorrelation function. Then* $\Delta_t(I)$ *increases with an increasing lag* $|t|$ This monotonicity property is the key idea for optimizing the value $\alpha$ at a given location. Instead of looking in the whole reference image for a the best match, an optimization algorithm is applied which saves us some computational time.

So far we have not developed an algorithm which would be optimal for the discrepancy minimization. However, even if the discrepancy norm can not be proven to be differentiable, its autocorrelation function is proven to have a Lipschitz behavior and is, according to Rademacher's theorem, smooth almost everywhere. Therefore the use of Newtonian optimization scheme and numerical differentiation seems to be a intelligent first guess.

### 3.3 Chosen Implementation

According to what have been said above, we have decided to combine two optimizations schemes: one we will be responsible for the global convergence heading to the basin of convergence, `Direct`[Shu72], and another one to define the local maximum based on a lagrangian method with bfgs[Bro70a, Bro70b, Fle70, Gol70, Sha70] updates.

The whole algorithm is described in algorithm 1.

**Algorithm 1** Basic OptLDMap

```
function OptLDMap(referenceImage, defectImage(size $M \times N$))
    input reference image
    for $x = 0, n, 2 * n, \cdots \div (N - n)$ do
        for $y = 0, m, 2 * m, \cdots \div (M - m)$ do
            set position of $m \times n$ sliding window to $(x, y)$ of defectImage
            repeat
                transform sliding window content with $H$ to referenceImage
                calculate discrepancy between sliding window and area in referenceImage
                recalculate transformation $H$ to minimize discrepancy
            until optimum with some $\epsilon$ is reached
            enter minimum distance in resultImage for the area of the sliding window
        end for
    end for
    binarize resultImage
    return resultImage
end function.
```

# 4 Experiments

In this section we are interested in showing some results obtained with our approach compared with the wavelet decomposition. We first describe the dataset we have been working on (in Section 4.1). Resulting images and remarks will be given un Section 4.2.

## 4.1 Dataset

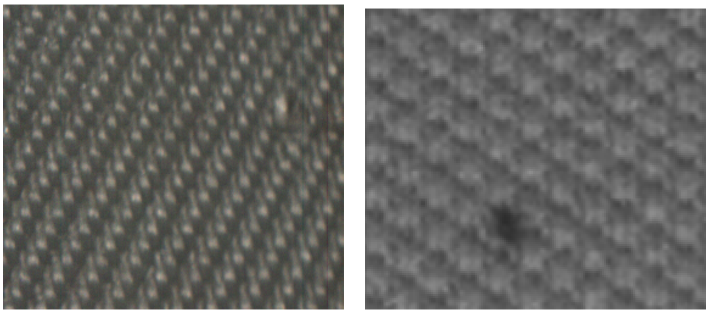
The textured images we are working on were taken from two independent sources. The first one comes from our industrial partner and examples are shown in Figure 2. The second source is the TILDA database. Some examples of this database are shown in Figure 3.

As we will see in the next Section the example 2(a) is considered to be a hard sample as its defect is rather smooth within the texture.

## 4.2 Results

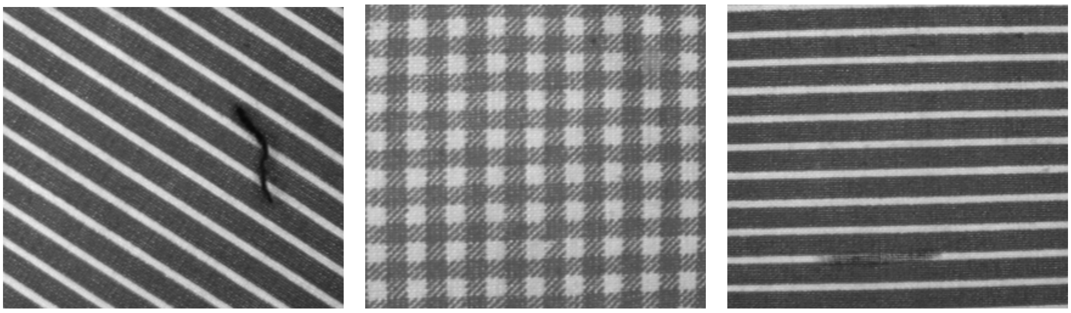
Here we want to show some visual results. We will compare our novel methods with one based on a Haar-wavelet decomposition. Visually, we can make our mind on Figure 4.

Different remarks can be stated. First, it appears that the 'hard sample' introduced in the previous section is indeed a challenging problem. The results obtained with our method shows that the algorithm needed to check for smaller details quite often as it can be seen on Figure 4(a). This defect is also hard to find for the wavelet based approach as shown on Figure 4(b). The defect on this image does not appear clearly and is not well localized.



(a) Airbag Hose (b) Woven textile

Figure 2: Some examples of defects in industrial textured surfaces.



(a) C3R3E4N23 (b) C3R1E2N3 (c) C3R3E2N39

Figure 3: Some examples of defects in the TILDA database.

On the examples shown, it seems that our algorithm manages better and more accurate results. In Figure 4(h) the results the map obtained by weighting the image with its defectiveness values shows not only the defective region (upper right corner) but other parts of the image too. In Figure 4(k), no defects are found. This could be due to the choice of the neighborhood area to calculate the features, or to the choice of the feature (energy in our experiments) itself. The wavelet decomposition shows anyway, that some work should be done to make use of the clear appearance of the defect in the approximation image (bottom right corner of the decomposition).

Moreover we would like to give some further remarks. Besides the comparable accuracy of each methods, one should remark that the wavelet based one is not easy to implement, because of its numerous (meta-)parameters to be set such as the classifier used (kNN, Neural network,...) the type of the wavelet, and the number of levels in the wavelet decomposition. Our novel algorithm only needs to be given a sliding window size.

As one can see, it is much harder to get a working algorithm based on the wavelet decomposition because of its hard parameter tuning. However, one has to agree that the implementation of our algorithm is still behind in terms of efficiency as it clearly is not working in real-time.

# 5 Conclusion

In this article we have reintroduced our approach for automatic inspection of regularly textured surfaces based on the discrepancy norm and template matching. We have extended the results obtained in our previous work to show its ability compared to a wavelet-based method.

So far our approach as shown promising results and seems to be adequate for the problem of texture inspection. However some work has to be done to make it efficient in comparison to other state of the art methods. Moreover more tests should be done in order to give the limitations (for instance in terms of frequency or noise) of our method.

One last point to be noted is the problem of optimizing the discrepancy norm during the registration process. We have not found at the moment a perfect optimized algorithm for the discrepancy minimization, but we are sure that the process we have used could be done faster by looking deeper into the minimization process.

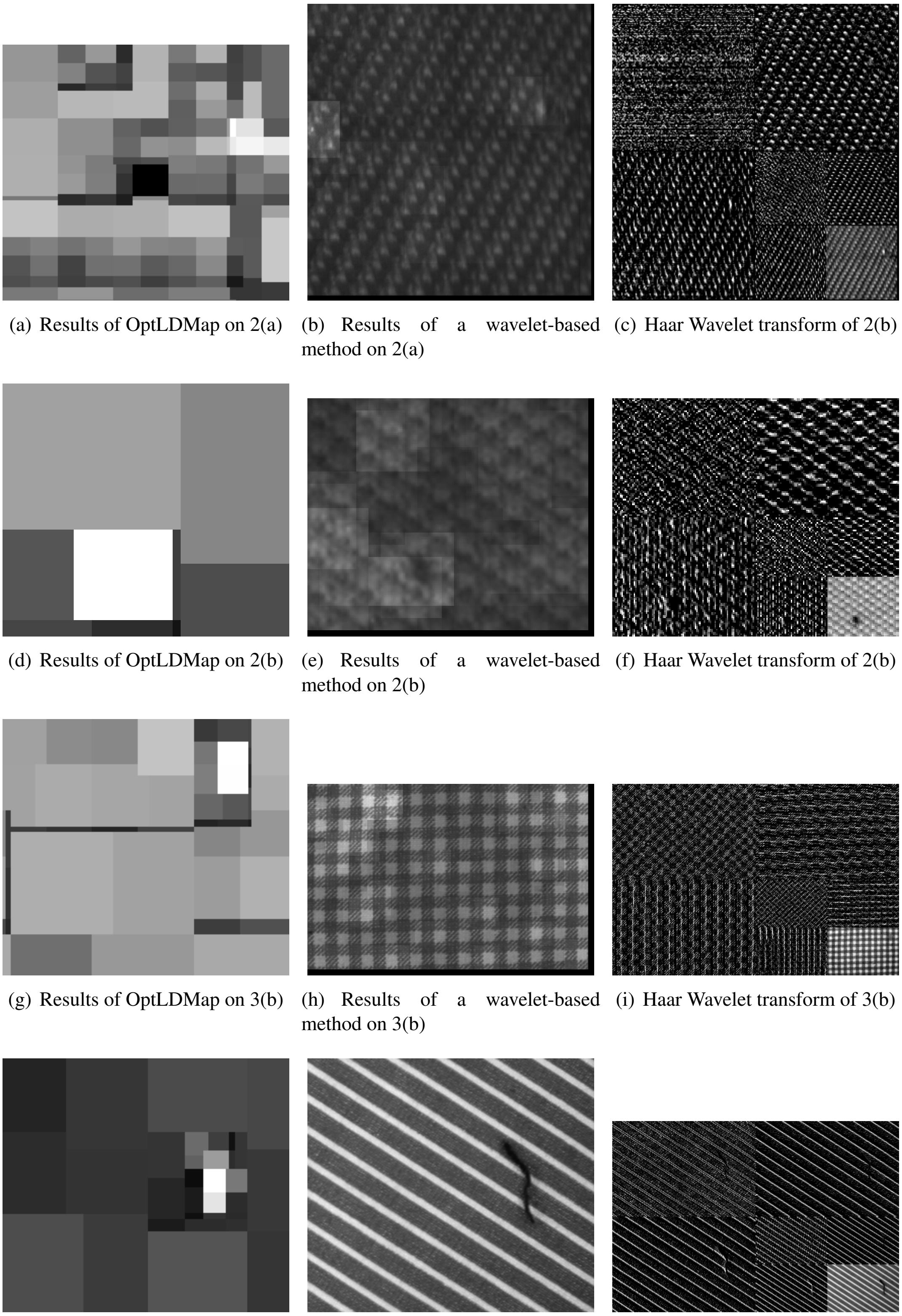
# References

[BBL02] A. Bodnarova, M. Bennamoun, and S. Latham. Optimal gabor filters for textile flaw detection. *Pattern Recognition*, 35:2973–2991, 2002.

[BNMR08] E. Baudrier, F. Nicolier, G. Millon, and S. Ruan. The adaptative local hausdorff-distance map as a new dissimilarity measure. *Pattern Recognition*, 41:1461–1478, 2008.

[Bro70a] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA J Appl Math*, 6(1):76–90, March 1970.

[Bro70b] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms: 2. The New Algorithm. *IMA J Appl Math*, 6(3):222–231, September 1970.



(a) Results of OptLDMap on 2(a)

(b) Results of a wavelet-based method on 2(a)

(c) Haar Wavelet transform of 2(b)

(d) Results of OptLDMap on 2(b)

(e) Results of a wavelet-based method on 2(b)

(f) Haar Wavelet transform of 2(b)

(g) Results of OptLDMap on 3(b)

(h) Results of a wavelet-based method on 3(b)

(i) Haar Wavelet transform of 3(b)

(j) Results of OptLDMap on 3(a)

(k) Results of a wavelet-based method on 3(a)

(l) Haar Wavelet transform of 3(a)

Figure 4: Results of two automatic defect detection algorithms on the images presented in 4.1. The first column shows the defectiveness map obtained with our method OptLDMap. The second one represents the defectiveness of the pixel based on the wavelet decomposition of the input image. In both columns, white means high defectiveness, and black means rather defect-free area. The third column shows the Haar-wavelet decomposition of the images. Some images have been resized to improve the speed of the algorithms.

[BSM11] Jean-Luc Bouchot, Gernot Stübl, and Bernhard Moser. A template matching approach based on the discrepancy norm for defect detection on regularly textured surfaces. accepted for presentation Quality Control by Automatic Vision 2011, June 2011.

[CP00] C. Costa and M. Petrou. Automatic registration of ceramic tiles for the purpose of fault detection. *Machine Vision and Applications*, 11(5):225–230, February 2000.

[Fle70] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, March 1970.

[Gol70] Donald Goldfarb. A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109):23–26, January 1970.

[HSD73] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *IEEE Trans. Syst., Man, Cybera*, 3:610–621, 1973.

[Kum08] Ajay Kumar. Computer-vision-based fabric defect detection: A survey. *IEEE Trans. Indus. Electr.*, 55:348–363, 2008.

[LB97] G. Lambert and F. Bock. Wavelet methods for texture defect detection. In *Image Processing, 1997. Proceedings., International Conference on*, volume 3, pages 201 – 204, Oct. 1997.

[LGALVGP01] Fernando López-García, Floreal Acebrón-Linuesa, José-Miguel Valiente-González, and Enrique Pérez. A study of registration methods for ceramic tiles inspection purposes. In *Proc. of IX Spanish Symp. on Pattern Recognition and Image Analysis*, pages 145–150, 2001.

[MMMJ09] M.Ghazvini, S. A. Monadjemi, N. Movahhedinia, and K. Jamshidi. Defect detection of tiles using 2d-wavelet transform and statistical features. In *WCSET 2009*, 2009.

[Mos09] B. Moser. Similarity measure for image and volumetric data based on Hermann Weyl's discrepancy measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 99:DOI 10.1109/TPAMI.2009.50, 2009.

[Ng07] Hui-Fuang Ng. Automatic thresholding for defect detection. *Pattern Recogn. Lett.*, 27:1644–1649, 2007.

[NPSN05] Henry Y.T. Ngan, Grantham K.H. Pang, S.P.Yung, and Michael K. Ng. Wavelet based methods on patterned fabric defect detection. *Pattern Recognition*, 38(4):559 – 576, 2005.

[NW87] H. Neunzert and B. Wetton. Pattern recognition using measure space metrics. Technical Report 28, Universität Kaiserslautern, Fachbereich Mathematik, November 1987.

[Sha70] D. F. Shanno. Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111):647–656, July 1970.

[Shu72] Bruno O. Shubert. A sequential method seeking the global maximum of a function. *SIAM Journal on Numerical Analysis*, 9(3):379–388, 1972.

[SSG96] Hamed Sari-Sarraf and James S. Goddard. On-line optical measurement and monitoring of yarn density in woven fabrics. In Frederick Y. Wu and Shenghua Ye, editors, *Autom. Opt. Inspection for Ind.: Theory, Technology, and Application*, pages 444–452. SPIE, November 1996.

[XG00] Pin Xie and Sheng-Uei Guan. A golden template self-generating method for patterned wafer inspection, 2000.

[Xie08] Xianghua Xie. A review of recent advances in surface defect detection using texture analysis techniques. *Electr. Letters on Computer Vision and Image Analysis*, 3:1–22, 2008.

# An axiomatic treatment of equality

Sándor Jenei*

Department of Knowledge-Based Mathematical Systems
Johannes Kepler University, Linz, Austria

## Abstract

Equality algebras were introduced in [3]. We will discuss the term equivalence of equivalential equality algebras and BCK-algebras with meet along with its consequences. Among them is a generalization of a result of Kabziński and Wroński ([4]), namely, we provide an equational characterization of equivalence operators of BCK-algebras with meet. Next, we shall investigate congruences and filters of equality algebras and prove that, quite surprisingly, the variety of equality algebras is a **1**-regular, arithmetic variety.

**Keywords:** Equality algebra, BCK-algebra, term equivalence, closure operator, **1**-regular, congruence permutable, congruence distributive variety

## References

[1] N. Galatos, P. Jipsen, T, Kowalski, H. Ono, Residuated Lattices: An Algebraic Glimpse at Substructural Logics, Volume 151. Studies in Logic and the Foundations of Mathematics (2007) pp. 532

[2] K. Iseki, S. Tanaka, An introduction to the theory of BCK-algebras, Math. Japon., 23 (1978),1–26.

[3] S. Jenei, Equality algebras, (submitted)

[4] J. Kabziński, A. Wroński, On equivalential algebras, Proceedings of the 1975 International Symposium on Multiple-valued Logic, Indiana University, Bloomington, 1975, pp. 231–243.

[5] M. Palasinski, On ideal and congruence lattices of BCK-algebras, Math., Japon., 26 (1981), 543–544.

*This work was supported by the OTKA-76811 grant and by the MC ERG grant 267589.

# The versatility of the Logic of Approximate Entailment

**Thomas Vetterlein**
Department of Knowledge-Based Mathematical Systems
Johannes Kepler University Linz
Altenbergerstraße 69, 4040 Linz, Austria
Thomas.Vetterlein@jku.at

Reasoning under practical circumstances is often inexact. Assumptions might be fulfilled only in an approximate way but conclusions are drawn anyway. In order to formalise such kind of reasoning classical propositional logic is not applicable. Our contribution deals instead with what is called similarity-based reasoning.

E. Ruspini initiated this issue in [Rus]. The topic has been worked out in a series of articles from a logical perspective; see, e.g., [Rod, GoRo]. In particular, the so-called *approximate entailment* was defined and studied.

The Logic of Approximate Entailment is based on the following idea. A set of worlds $W$ is endowed with a similarity relation $s$. The function $s$ assigns to each pair of worlds $v, w \in W$ a real degree between 0 and 1; $s(v, w)$ is the larger the more $v$ resembles $w$. Properties $\varphi$ and $\psi$ are modelled classically: by subsets $A$ and $B$ of $W$, respectively. Given a degree $d \in [0, 1]$, an implication of the form

$$\varphi \overset{d}{\Rightarrow} \psi$$

is satisfied if

$$A \subseteq U_d(B)$$

holds. Here, $U_d(B)$ is the $d$-neighborhood of $B$ in $W$.

We deal with the axiomatics of this logic and point out both solutions and challenges. We furthermore underline the generality and consequently the wide applicability of the idea, in particular in connection with reasoning under vagueness.

## References

[GoRo] L. Godo, R. O. Rodríguez. Logical approaches to fuzzy similarity-based reasoning: an overview, in: G. Della Riccia et al. (eds.). *Preferences and similarities*, CISM Courses and Lectures 504, Springer-Verlag, Berlin 2008; pp. 75 - 128.

[Rod] R. O. Rodríguez. *Aspectos formales en el Razonamiento basado en Relaciones de Similitud Borrosas*, *Ph. D. Thesis*, Technical University of Catalonia (UPC), 2002.

[Rus] E. Ruspini. On the semantics of fuzzy logic, *International Journal of Approximate Reasoning* **5** (1991), 45 - 88.