Chemometrics and Intelligent Laboratory Systems 96 (2009) 182–187

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Confidence interval of fuzzy models: An example using a waste-water treatment plant

Igor Škrjanc *

Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 2 April 2008
Received in revised form 26 January 2009
Accepted 29 January 2009
Available online 10 February 2009

Keywords:
Fuzzy model
Interval fuzzy model
Confidence interval

ABSTRACT

In this paper we present a new approach to fuzzy confidence interval identification. The method combines a fuzzy identification methodology with some ideas from applied statistics. The idea is to find, on a finite set of measured data, the confidence interval defined by the lower and upper fuzzy bound that define the band that contains all the output measurements. The method can be successfully used when we are trying to describe a family of uncertain nonlinear functions or when we are trying to find the interval for a nonlinear process output where all the measurements can be found. The fuzzy confidence interval model can be used in process monitoring, fault detection or in the case of robust control design. In our example the proposed method is used for waste-water treatment plant modeling, which exhibit a very nonlinear behavior.

1. Introduction

The problem of nonlinear model identification from a finite set of measured data using an optimality criterion has received a great deal of attention in the scientific community. Many different approaches have appeared to approximate functions from data, these include the following: the continuous piecewise linear (PWL) approach, where it is possible to uniformly approximate any Lipschitz continuous function defined on a compact domain using linear functions [7]; the neural-network approach [9], which is universal approximation, but has a drawback in the sense of the interpretability of the result; and the fuzzy model approach, which in Takagi–Sugeno (TS) form, approximates the nonlinear system by smoothly interpolating affine local models [11] with, each local model contributing to the global model in a fuzzy subset of the space characterized by a membership function.

In this paper we look at the development of a confidence interval approximation based on a fuzzy model. This results in a lower and an upper fuzzy bounds. The interval fuzzy model identification is a methodology for approximating the nonlinear functions of a finite set of input and output measurements that can also be used to compress the information in the case of an approximation of a nonlinear function family to obtain the interval or the band containing the whole set of measurements. The interval fuzzy model approach is described in [4] and [10], where the linear programming approach is used to obtain the fuzzy confidence interval, and in [6], where the confidence band is obtained using a least-squares optimization and a constant variance is assumed across the whole problem domain. In our approach the variance of the noise can vary, i.e. the method can cope with a heteroskedastic noise.

The proposed method is of great importance in many technological areas, e.g., the modeling of nonlinear time-invariant systems with uncertain physical parameters. In our example the method of confidence interval modeling is used to model the behavior of a waste-water treatment plant. These types of plants are, due to their nature, subjected to daily, weekly and seasonal variations because of temperature changes, rain and a varying process load. In the case of a waste-water treatment plant the theoretical modeling is a very demanding task with questionable results. For this reason the methods of data mining have been adopted. However, we were not able to find a nonlinear model which will approximate the behavior of the plant for the whole set of measured data i.e., with a sufficiently small error, but we could find the confidence interval in which we always find the output variable.

The paper is organized as follows: Section 2 provides the background to the fuzzy modeling; Section 3 describes the idea of fuzzy confidence interval model identification; Section 4 introduces the confidence interval of the local linear model; and Section 5 presents an application of the confidence interval modeling.

2. Nonlinear model described in fuzzy form

A typical fuzzy model [11] is given in the form of rules

$$\begin{aligned} \mathbf{R}_{j_1,\ldots,j_q}: \ & \text{if } x_{p1} \text{ is } \mathbf{A}_{1,j_1} \text{ and } x_{p2} \text{ is } \mathbf{A}_{2,j_2} \text{ and } \ldots \text{ and } x_{pq} \text{ is } \mathbf{A}_{q,j_q} \text{ then } y = \phi_{j_1,\ldots,j_q}(\boldsymbol{x}) \\ & j_1 = 1,\ldots,m_1 \quad j_2 = 1,\ldots,m_2 \quad \ldots \quad j_q = 1,\ldots,m_q \end{aligned} \tag{1}$$

The q-element vector $\boldsymbol{x}_p^T = [x_{p1},\ldots,x_{pq}]$ denotes the input or variables in premise, and the variable y is the output of the model. With each variable in premise x_{pi} $(i=1,\ldots,q)$, f_i is connected to the fuzzy sets ($\mathbf{A}_{i,1}$, …, $\mathbf{A}_{i,m_i}$), and each fuzzy set $\mathbf{A}_{i,j_i}$ $(j_i=1,\ldots,m_i)$ is associated with a real-valued function $\mu_{A_{i,j_i}}(x_{pi}) : \mathbb{R}\rightarrow[0,1]$ that produces the membership

* Tel.: +386 1 4768 31; fax: +386 1 4264 631.
E-mail address: igor.skrjanc@fe.uni-lj.si.

0169-7439/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2009.01.009

A list of used symbols

$\mathbf{R}_{j1,\dots,jq}$, $\mathbf{R}_j$,	the fuzzy rule
$\boldsymbol{A}_{i,j}$,	the jth fuzzy set of ith variable in the rule premise
$\phi_{j1,\dots,jq}$, ϕ_j	the function in the fuzzy rule consequence
$\beta_j(\boldsymbol{x}_p)$	the degree of fulfillment for jth rule
$\boldsymbol{S}$	the set of measured data
$\boldsymbol{X}_p$	the set of data in the rule premise
$\boldsymbol{X}$	the set of data in the rule consequence
$\boldsymbol{Y}$	the set of output data
$\boldsymbol{x}_p$	the vector of variables in the premise part of the rule
$\boldsymbol{x}$	the vector of variables in the consequence part of the rule
y_i	the output at time instant i
$y_{i,j}$	the output of the jth local linear model at time instant i
$\boldsymbol{\psi}_j$	the input vector $\boldsymbol{x}$ weighted by the degree of fulfillment $\beta_j(\boldsymbol{x}_p)$
$\boldsymbol{\theta}_j$	the set of coefficient of jth linear model
$\boldsymbol{\Theta}$	the matrix of coefficients of all local linear models
e_i	the error between the output measurement and the estimated output value
$e_{i,j}$	the error between the weighted output measurement and the estimated output of jth local linear model
$\underline{f}, \overline{f}$	the lower and upper fuzzy function

grade of the variable x_{pi} with respect to the fuzzy set $\mathbf{A}_{i,j_i}$. To make the list of fuzzy rules complete, all possible variations of the fuzzy sets given in Eq. (1) are required. This gives the number of fuzzy rules $m = m_1 \times m_2 \times \ldots \times m_q$. The variables x_{pi} are not the only inputs of the fuzzy system. Implicitly, the $n+1$-element vector $\boldsymbol{x}^T = [1, x_1, \dots, x_n]$ also represents an input to the system. It is usually referred to as the consequence vector. The functions $\phi_{j_1,\dots,jq}(\cdot)$ can be arbitrary smooth functions in general, although linear or affine functions are normally used. This means that the rule $\mathbf{R}_{j_1,\dots,jq}$ can consequently also be denoted as $\mathbf{R}_j$, where j means

$$j = j_1 + (j_2 - 1)\Pi_{i=1}^{1} m_i + (j_3 - 1)\Pi_{i=1}^{2} m_i + \ldots + \left(j_q - 1\right)\Pi_{i=1}^{q-1} m_i,$$

where $j_1 = 1, \dots, m_1$, $j_2 = 1, \dots, m_2, \dots, j_q = 1, \dots, m_q$. The same notation can also be used for the function $\phi_{j_1,\dots,jq}(\cdot)$.

The fulfillment of the rule $\mathbf{R}_j$, i.e., the rule $\mathbf{R}_{j_1,\dots,jq}$ means the fulfillment of the premise of the rule in a normalized form is then expressed with the functions $\beta_j(x_p)$, defined as

$$\beta_j\left(\boldsymbol{x}_p\right) = \frac{\mu_{A_{1,j_1}}\left(x_{p1}\right)\mu_{A_{2,j_2}}\left(x_{p2}\right)\dots\mu_{A_{q,jq}}\left(x_{pq}\right)}{\sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2}\dots\sum_{j_q=1}^{m_q}\mu_{A_{1,j_1}}\left(x_{p1}\right)\mu_{A_{2,j_2}}\left(x_{p2}\right)\dots\mu_{A_{q,jq}}\left(x_{pq}\right)}, \quad (2)$$

The fulfillment of the rule $\mathbf{R}_j$, denoted as $\beta_j(\boldsymbol{x}_p)$, gives information about the participation of the function $\phi_j(\cdot)$ in the whole output variable. The output variable y is then given as a weighted sum defined as

$$y = \sum_{j=1}^{m} \beta_j\left(\boldsymbol{x}_p\right)\phi_j(\boldsymbol{x}) \quad (3)$$

A very frequently used structure of the fuzzy model known from the literature [9] is the structure with the output value defined as a linear combination of the consequence states

$$\phi_j(\boldsymbol{x}) = \boldsymbol{x}^T\theta_j, \quad \theta_j^T = \left[\theta_{j0}, \theta_{j1}, \dots, \theta_{jn}\right]^T, \quad j = 1, \dots, m \quad (4)$$

In this case, Eq. (3) consists of m local linear models and can be written as

$$y = \sum_{j=1}^{m} \psi_j^T\theta_j, \quad j = 1, \dots, m, \quad (5)$$

where

$$\psi_j^T = \beta_j\left(\boldsymbol{x}_p\right)\boldsymbol{x}^T \quad (6)$$

We assume a set of measured data defined as the set $\boldsymbol{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_N\}$, where $\boldsymbol{s}_i$ stands for a vector of all the measurements at the time instant i. From this set the premise vectors $\boldsymbol{X}_p = \{\boldsymbol{x}_{p1}, \boldsymbol{x}_{p2}, \dots, \boldsymbol{x}_{pN}\}$ and the set of antecedent (or consequence) vectors $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\}$ are constructed by choosing a certain measurement to be a part of $\boldsymbol{X}_p$ or $\boldsymbol{X}$. Eq. (6) shows that the output of a fuzzy system is a function of the premise vector $\boldsymbol{x}_p$ (q-dimensional) and the consequence vector $\boldsymbol{x}$ (n-dimensional). Therefore, the data set $\boldsymbol{Z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_N\}$, where $\boldsymbol{z}_i^T = [\boldsymbol{x}_{pi}^T, \boldsymbol{x}_i^T]$, $i = 1, \dots, N$, is constructed to simplify the notation in Eq. (6), as follows

$$y = \sum_{j=1}^{m} \psi_j^T(\boldsymbol{z})\theta_j, \quad j = 1, \dots, m, \quad (7)$$

where $\boldsymbol{\psi}^T{}_j(\boldsymbol{z})$ denotes the dependence on both the input vectors $\boldsymbol{x}$ and $\boldsymbol{x}_p$.

If the matrix of the coefficients for the whole set of rules is written as $\boldsymbol{\Theta}^T = [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_m^T]$, and the fuzzy regression matrix

$$\psi^T = \left[\psi_1^T, \dots, \psi_m^T\right] \quad (8)$$

then Eq. (3) can be rewritten in the matrix form

$$y = \psi^T(\boldsymbol{z})\boldsymbol{\Theta} \quad (9)$$

The fuzzy model in the form given in Eq. (9) is referred to as the affine Takagi–Sugeno model and can be used to approximate any arbitrary function that maps the compact set of inputs to the compact set of outputs with any desired degree of accuracy [8,12,13]. This generality can be proven with the Stone–Weierstrass theorem [5], which suggests that any continuous function can be approximated by a fuzzy basis function expansion [9]. The parameters of the fuzzy model $\boldsymbol{\Theta}$ are obtained with the least-squares method [11].

3. Fuzzy confidence interval identification

The fuzzy model structure will be used to define the band of lower and upper bounds, which is called the fuzzy confidence interval. The basic idea is to find, from a finite set of measured data, the confidence interval that contains all the output measurements. In the same way as the sets $\boldsymbol{X}_p$ and $\boldsymbol{X}$, a set of corresponding outputs is also defined from the set of measured data $\boldsymbol{S}$ as $\boldsymbol{Y} = \{y_1, y_2, \dots, y_N\}$ and is the result of mapping the input data set $\boldsymbol{Z}$ on the nonlinear real continuous function g. This mapping can be written as follows

$$y_i = g(\boldsymbol{z}_i), \quad i = 1, \dots, N \quad (10)$$

According to the Stone–Weierstrass theorem, for any given real continuous function g and arbitrary $\varepsilon > 0$, there exists a fuzzy system such that

$$\max_i |y_i - \psi^T(\boldsymbol{z}_i)\boldsymbol{\Theta}| < \epsilon, \quad (11)$$

where $\boldsymbol{\psi}^T(z_i)\boldsymbol{\Theta}$ stands for the output of the fuzzy model for a certain measured input vector $\boldsymbol{z}_i$. This implies the approximation of any given real continuous function with the fuzzy function defined in Eq. (9). However, it should be pointed out that the lower values of ϵ imply higher values of the required rules m that satisfy Eq. (11). The answer lies in the proper arrangement of the membership functions. This is a well-known problem in fuzzy systems. However, it can be overcome with a cluster analysis [2,3] or a hierarchical division of the input space, or other approaches. The details will not be discussed in this paper.

Taking into account Eq. (9) and Eq. (11) the set of data samples can be written as follows:

$$y_i = \psi^T(\boldsymbol{z}_i)\boldsymbol{\Theta} + e_i, \quad i = 1, \ldots, N \tag{12}$$

where e_i, $i=1,\ldots, N$, stands for the heteroskedastic noise of zero mean value, which is written as $\boldsymbol{e}=N(0, \sigma^2(\boldsymbol{x}_p))$, where $\boldsymbol{e}=[e_1,\ldots, e_N]^T$.

The error between the measured values and the fuzzy function outputs can be defined as

$$e_i = y_i - \psi^T(\boldsymbol{z}_i)\boldsymbol{\Theta}, \quad i = 1, \ldots, N \tag{13}$$

By having the membership functions defined, the structure of the model is known and only the fuzzy model parameters have to be defined. The parameters of $\boldsymbol{\Theta}$ are calculated separately for each local model. This means that we split Eq. (13) into m equations of the form

$$e_{i,j} = y_{i,j} - \psi_{i,j}^T\theta_j, \quad i = 1, \ldots, N, \quad j = 1, \ldots, m \tag{14}$$

where $e_{i,j}=\beta_{i,j}e_i$, $y_{i,j}=\beta_{i,j}y_i$, $\boldsymbol{\psi}_{i,j}^T=\beta_{i,j}\boldsymbol{\psi}^T(\boldsymbol{z}_i)$ and where $\beta_{i,j}=\beta_j(\boldsymbol{x}_{pi})$, $i=1,\ldots, N$. In matrix form the equation is written as follows

$$\boldsymbol{e}_j = \boldsymbol{y}_j - \boldsymbol{\Psi}_j^T\theta_j, \quad j = 1, \ldots, m \tag{15}$$

where $\boldsymbol{e}_j=[e_{1,j},\ldots, e_{N,j}]^T$, $\boldsymbol{y}_j=[y_{1,j},\ldots, y_{N,j}]^T$ and $\Psi_j=[\boldsymbol{\psi}_{1,j},\ldots, \boldsymbol{\psi}_{N,j}]^T$.

The vector of the estimated local model parameters is the minimizing argument, which can be expressed as

$$\hat{\theta}_j = \arg\min V_j\left(\theta_j\right), \quad j = 1, \ldots, m \tag{16}$$

where V_j reads as $V_j=\boldsymbol{e}_j^T\boldsymbol{e}_j$. The idea of an approximation can be interpreted as the most representative local fuzzy function to describe the local domain of the outputs $\boldsymbol{y}_j$ as a function of the inputs $\boldsymbol{z}$. The estimation of the local fuzzy model parameters is given by the *minimum least-squares* optimization, as follows:

$$\hat{\theta}_j = \left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\boldsymbol{\Psi}_j\boldsymbol{y}_j \tag{17}$$

and the estimated output of the jth local fuzzy model is therefore written as $\hat{\boldsymbol{y}}_j=\boldsymbol{\psi}_j^T\hat{\boldsymbol{\theta}}_j$. In a particular case the estimated parameters of the fuzzy model, taking into account Eq. (15), become $\hat{\boldsymbol{\theta}}_j=\boldsymbol{\theta}_j+\tilde{\boldsymbol{\theta}}_j$, where

$$\tilde{\theta}_j = \left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\boldsymbol{\Psi}_j\boldsymbol{e}_j \tag{18}$$

The expected bias of the local model parameters is then described as follows:

$$E\left\{\hat{\theta}_j\right\} = \theta_j + E\left\{\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\boldsymbol{\Psi}_j\boldsymbol{e}_j\right\}, \quad j = 1, \ldots, m \tag{19}$$

The right term in Eq. (19) equals zero, because of the uncorrelated regression matrix $\boldsymbol{\Psi}_j$, the vector $\boldsymbol{e}_j$, and the zero mean value $E\{\boldsymbol{e}_j\}=0$. This can be explained by taking into account the statistical property of the noise $E\{\boldsymbol{e}\}=0$, which implies that the noise of the jth local linear model, i.e., the weighted mean value $E\{\boldsymbol{e}_j\}=0$, also equals zero, when assuming a sufficiently large number of measurements inside a single fuzzy partitioning. This means that the estimation of the model parameters is unbiased. The weighted mean value is calculated as follows:

$$\bar{e}_j = E\left\{\boldsymbol{e}_j\right\} = \frac{1}{v}\sum_{i=1}^{N}\beta_{i,j}e_i, \quad v = \sum_{i=1}^{N}\beta_{i,j}, \quad j = 1, \ldots, m \tag{20}$$

The expected covariance of the estimated parameters is calculated in the following way:

$$\operatorname{cov}\left(\theta_j - \hat{\theta}_j\right) = E\left\{\tilde{\theta}_j\tilde{\theta}_j^T\right\} \tag{21}$$

taking into account that $E\{\tilde{\boldsymbol{\theta}}_j\}=0$. Using Eq. (18), the covariance matrix of model parameters is written as follows

$$\operatorname{cov}\left(\theta_j - \hat{\theta}_j\right) = E\left\{\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\boldsymbol{\Psi}_j\boldsymbol{e}_j\boldsymbol{e}_j^T\boldsymbol{\Psi}_j^T\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\right\} = \left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\boldsymbol{\Psi}_jE\left\{\boldsymbol{e}_j\boldsymbol{e}_j^T\right\}\boldsymbol{\Psi}_j^T\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1} \tag{22}$$

and by taking into account the following notation $E\{\boldsymbol{e}_j\boldsymbol{e}_j^T\}=\hat{\sigma}_j^2\boldsymbol{I}$ the covariance matrix is written as follows:

$$\operatorname{cov}\left(\theta_j - \hat{\theta}_j\right) = \hat{\sigma}_j{}^2\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1} \tag{23}$$

where $\hat{\sigma}_j^2$ stands for the weighted variance of $\boldsymbol{e}_j$

$$\hat{\sigma}_j{}^2 = \frac{1}{\mu - (n+1)}\sum\nolimits_{i=1}^{N}\beta_{i,j}^2\left(e_i - \bar{e}_j\right)^2, \quad \mu = \sum_{i=1}^{N}\beta_{i,j}^2, \quad j = 1, \ldots, m \tag{24}$$

and $n+1$ stands for the number of estimated parameters of the fuzzy model.

The expected covariance of the residuals between the observed data and the model output is given as follows

$$\operatorname{cov}\left(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j\right) = E\left\{\left(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j - E\left\{\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j\right\}\right)\left(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j - E\left\{\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j\right\}\right)^T\right\}. \tag{25}$$

Taking into account that $E\{\boldsymbol{e}_j\}=0$, the expected value of the residue between the measured output and the estimated output becomes $E\{\boldsymbol{y}_j-\hat{\boldsymbol{y}}_j\}=0$.

The covariance matrix of the residuals in Eq. (25) can be written as:

$$\operatorname{cov}\left(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j\right) = E\left\{\left(\boldsymbol{e}_j - \boldsymbol{\Psi}_j^T\tilde{\theta}_j\right)\left(\boldsymbol{e}_j - \boldsymbol{\Psi}_j^T\tilde{\theta}_j\right)^T\right\} \tag{26}$$

and by taking into account Eq. (18), it follows that

$$\operatorname{cov}\left(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j\right) = \hat{\sigma}_j{}^2\boldsymbol{I} - \hat{\sigma}_j{}^2\boldsymbol{\Psi}_j^T\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\boldsymbol{\Psi}_j. \tag{27}$$

3.1. Confidence interval definition

Let us define a confidence interval for the new set of data, generated by the same function g as in the case of the identification. The corresponding set of measured output values $Y^*=\{y_1^*,\ldots, y_M^*\}$ over the set of measured inputs $\boldsymbol{Z}^*$, i.e., $y_i^*=g(\boldsymbol{z}_i^*)$, $i=1,\ldots, M$, is called the validation data set.

The idea of confidence interval fuzzy modeling is to find a lower fuzzy output $\underline{f}$ and an upper fuzzy function $\bar{f}$ satisfying

$$\underline{f}\left(\boldsymbol{z}_i^*\right) \le g\left(\boldsymbol{z}_i^*\right) \le \bar{f}\left(\boldsymbol{z}_i^*\right), \quad \forall \boldsymbol{z}_i^*. \tag{28}$$

The main requirement when defining the band is that it is as narrow as possible and should contain a certain percentage of the data. This problem has been treated in the literature using the piecewise linear function approximation [7]. Our approach, using the fuzzy

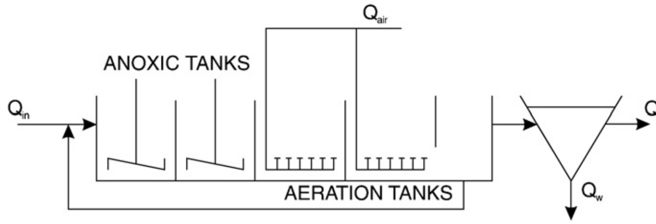

Fig. 1. Schematic representation of simulation benchmark.

function approximation, can be viewed as a generalization of the piecewise linear approach and gives a better approximation, or at least a much narrower approximation band.

The measured output values of the jth local linear model are now defined as

$$\boldsymbol{y}_j^* = \boldsymbol{\Psi}_j^{*T}\theta_j + \boldsymbol{e}_j^* \tag{29}$$

where $\boldsymbol{\Psi}_j^{*T}$ stands for the regression matrix of the jth local linear model and $\boldsymbol{y}_j^* = [y_1^*,\ldots,y_M^*]^T$. The model output of the jth local linear model is, in the case of the validation data set, defined as follows:

$$\hat{\boldsymbol{y}}_j^* = \boldsymbol{\Psi}_j^{*T}\hat{\theta}_j. \tag{30}$$

To calculate the confidence interval in the case of a validation set, we have to calculate the expected covariance of the residual between the model output and the new set of data in each local domain

$$\operatorname{cov}\left(\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^*\right) = E\left\{\left(\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^* - E\left\{\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^*\right\}\right)\left(\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^* - E\left\{\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^*\right\}\right)^T\right\}. \tag{31}$$

Taking into account the same statistical properties of the noise for the data in the validation data set ($E\{\boldsymbol{e}_j^*\}=0$) and for the identification set ($E\{\boldsymbol{e}_j\}=0$), the expected value of the error between the measured output and the estimated output becomes $E\{\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^*\}=0$.

The covariance matrix in Eq. (31) can be rewritten as:

$$\operatorname{cov}\left(\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^*\right) = E\left\{\left(\boldsymbol{e}_j^* - \boldsymbol{\Psi}_j^{*T}\tilde{\theta}_j\right)\left(\boldsymbol{e}_j^* - \boldsymbol{\Psi}_j^{*T}\tilde{\theta}_j\right)^T\right\} \tag{32}$$

and subsequently as follows:

$$\operatorname{cov}\left(\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^*\right) = E\left\{\boldsymbol{e}_j^*\boldsymbol{e}_j^{*T}\right\} - E\left\{\boldsymbol{\Psi}_j^{*T}\tilde{\theta}_j\boldsymbol{e}_j^{*T}\right\} - E\left\{\boldsymbol{e}_j^*\tilde{\theta}_j^T\boldsymbol{\Psi}_j^*\right\} + E\left\{\boldsymbol{\Psi}_j^{*T}\tilde{\theta}_j\tilde{\theta}_j^T\boldsymbol{\Psi}_j^*\right\}. \tag{33}$$

Taking into account Eq. (18) and assuming that both the noise signals have identical statistical properties, $E\{\boldsymbol{e}_j\boldsymbol{e}_j^T\} = E\{\boldsymbol{e}_j^*\boldsymbol{e}_j^{*T}\} = \hat{\sigma}_j^2$, and are uncorrelated $E\{\boldsymbol{e}_j\boldsymbol{e}_j^{*T}\} = E\{\boldsymbol{e}_j^*\boldsymbol{e}_j^T\} = 0$, Eq. (33) is written as follows

$$\operatorname{cov}\left(\boldsymbol{y}_j^* - \hat{\boldsymbol{y}}_j^*\right) = \hat{\sigma}_j^2\boldsymbol{I} + \hat{\sigma}_j^2\boldsymbol{\Psi}_j^{*T}\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\boldsymbol{\Psi}_j^* \tag{34}$$

The lower and the upper confidence intervals of the local linear model are therefore defined as

$$\underline{f}_j\left(\boldsymbol{z}_i^*\right) = \psi_{i,j}^{*T}\theta_j - t_{\alpha,M-n}\hat{\sigma}_j\left(1 + \psi_{i,j}^{*T}\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\psi_{i,j}^*\right)^{\frac{1}{2}}, \quad i = 1,\ldots,M \tag{35}$$

and

$$\bar{f}_j\left(\boldsymbol{z}_i^*\right) = \psi_{i,j}^{*T}\theta_j + t_{\alpha,M-n}\hat{\sigma}_j\left(1 + \psi_{i,j}^{*T}\left(\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^T\right)^{-1}\psi_{i,j}^*\right)^{\frac{1}{2}}, \quad i = 1,\ldots,M \tag{36}$$

where $t_{\alpha,M-n}$ stands for the percentile of the t-distribution for $100(1-2\alpha)$ percentage confidence interval with $M-n$ degrees of freedom.

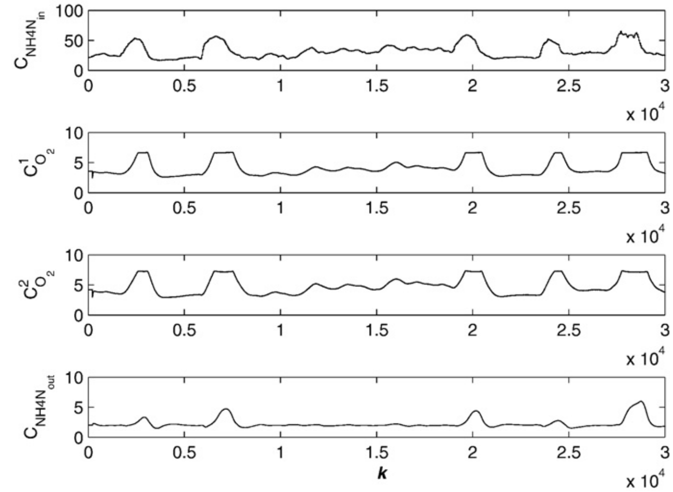

Fig. 2. The whole set of measurements.

The lower and upper fuzzy functions are then described as:

$$\begin{aligned} \underline{f}\left(\boldsymbol{z}_i^*\right) &= \sum_{i=1}^{m}\underline{f}_j\left(\boldsymbol{z}_i^*\right) \\ \bar{f}\left(\boldsymbol{z}_i^*\right) &= \sum_{i=1}^{m}\bar{f}_j\left(\boldsymbol{z}_i^*\right) \end{aligned} \tag{37}$$

Interval fuzzy modeling can be used efficiently in the case of fault detection, where the data set of normal operating systems is modeled by the interval fuzzy model to obtain the band of normal functioning. During operations this band is calculated online and is checked to see if a measurement corresponds to the normal functioning band or not. If the measurement violates the tolerance band, one can assume that a malfunction might have occurred. The proposed model can also be used for the case of robust control design, as described in [1].

4. Confidence interval for biological waste-water treatment process

Waste-water treatment plants are large, nonlinear systems subject to large perturbations in flow and load, together with uncertainties concerning the composition of the incoming waste water. The simulation benchmark has been developed to provide an unbiased system

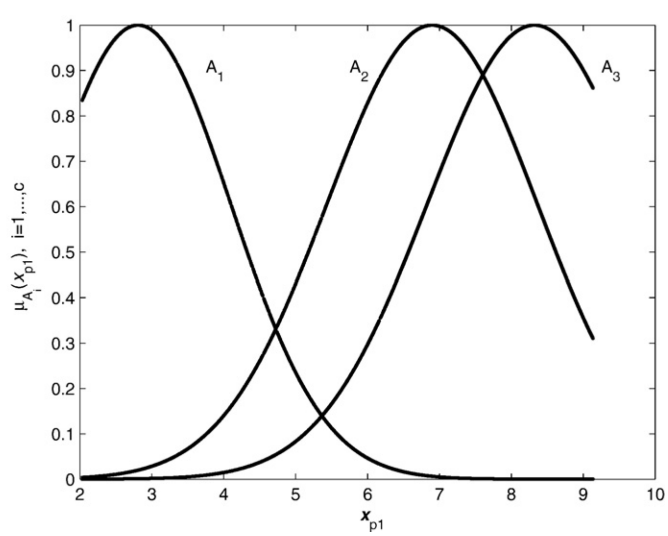

Fig. 3. The distribution of membership functions.

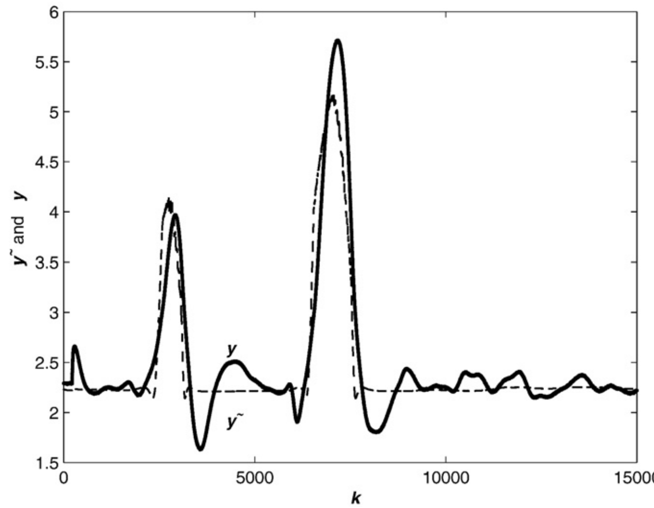

Fig. 4. The verification of fuzzy model.

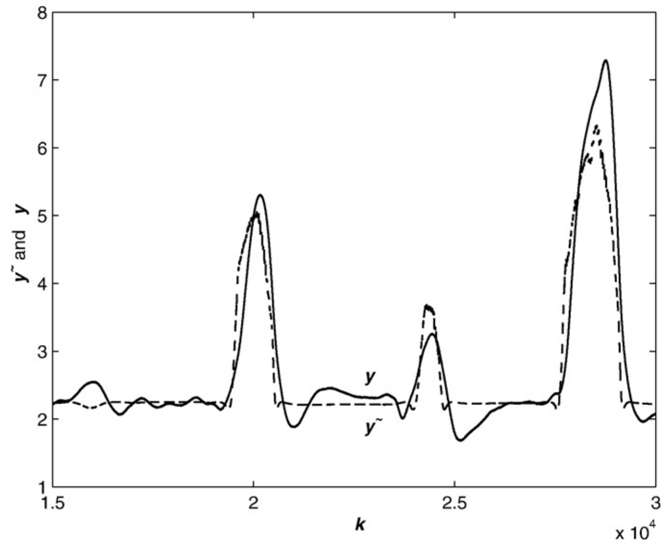

Fig. 6. The validation of fuzzy model.

for comparing various strategies without reference to a particular facility. It consists of five sequentially connected reactors along with a 10-layer secondary settling tank. The plant layout, model equations and control strategy are described in detail on a *www* page (http://www.ensic.u-nancy.fr/costwwtp). In our approach the layout was formed such that the waste water is purified in the mechanical phase, and after this phase the moving bed bio-film reactor is used. A schematic representation of the simulation benchmark is shown in Fig. 1.

The confidence interval was constructed for the simulation model, where the following measurements were used: the influent ammonia concentration in the inflow Q_{in}, defined as $C_{\text{NH4N}_{\text{in}}}$; the dissolved oxygen concentration in the first aerobic reactor tank $C^1_{\text{O}_2}$; the dissolved oxygen concentration in the second aerobic reactor tank $C^2_{\text{O}_2}$; and the ammonia concentration in the second aerobic reactor tank $C_{\text{NH4N}_{\text{out}}}$. All the measurements define the input data set $\boldsymbol{S}=\{C_{\text{NH4N}_{in}}(k), C^1_{\text{O}_2}(k), C^2_{\text{O}_2}(k), C_{\text{NH4N}_{\text{out}}}(k)\}$, $k=1,\ldots N$. The fuzzy model was built to model the relation between the ammonia concentration in the second aerobic reactor tank and the other measured variables:

$$C_{\text{NH4N}_{\text{out}}}(k+1) = g\left(C_{\text{NH4N}_{\text{in}}}(k), C^1_{\text{O}_2}(k), C^2_{\text{O}_2}(k), C_{\text{NH4N}_{\text{out}}}(k)\right) \tag{38}$$

where g stands for the nonlinear relation between the measured variables. The variables are measured with the sampling time $T_s=120$ s. The whole set of measurements is shown in Fig. 2.

The structure of the fuzzy model, which is obtained on the set of the first 15,000 samples, is as follows taking into account Eq. (1)

$$\text{R}_j: \text{ if } x_{p1}(k) \text{ is } \mathbf{A}_j \text{ then } y(k) = \phi_j(\boldsymbol{x}(k)), \quad j=1,2,3 \tag{39}$$

where x_{p1} stands for $C_{\text{NH4N}_{\text{out}}}(k)$, $\boldsymbol{x}^T(k)=[C_{\text{NH4N}_{\text{in}}}(k), C^1_{\text{O}_2}(k)(k), C^2_{\text{O}_2}(k), C_{\text{NH4N}_{\text{out}}}(k)]$ and $y(k)$ is equal to $C_{\text{NH4N}_{\text{out}}}(k+1)$. Taking into account a local linear function ϕ_j the following structure is obtained

$$\begin{aligned}&\text{R}_j: \text{ if } C_{\text{NH4N}_{\text{out}}}(k) \text{ is } \mathbf{A}_j \text{ then}\\ &C_{\text{NH4N}_{\text{out}}}(k+1) = \theta_{j0} + \theta_{j1}C_{\text{NH4N}_{\text{in}}}(k) + \theta_{j2}C^1_{\text{O}_2}(k) + \theta_{j3}C^2_{\text{O}_2}(k) + \theta_{j4}C_{\text{NH4N}_{\text{out}}}(k),\\ &j=1,2,3.\end{aligned} \tag{40}$$

The input data set was then clustered using the Gustafson–Kessel fuzzy clustering algorithm to divide the domain into fuzzy subspaces. In our example the number of clusters was chosen by the trial-and-fail method ($c=3$). The membership functions are shown in Fig. 3.

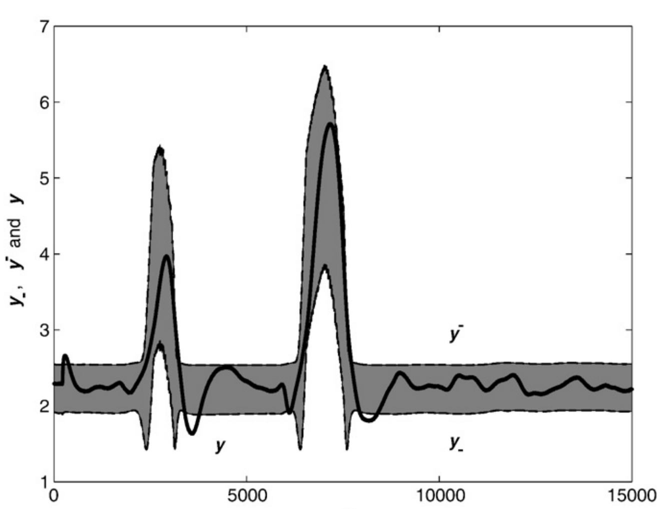

Fig. 5. The verification of fuzzy model with confidence interval.

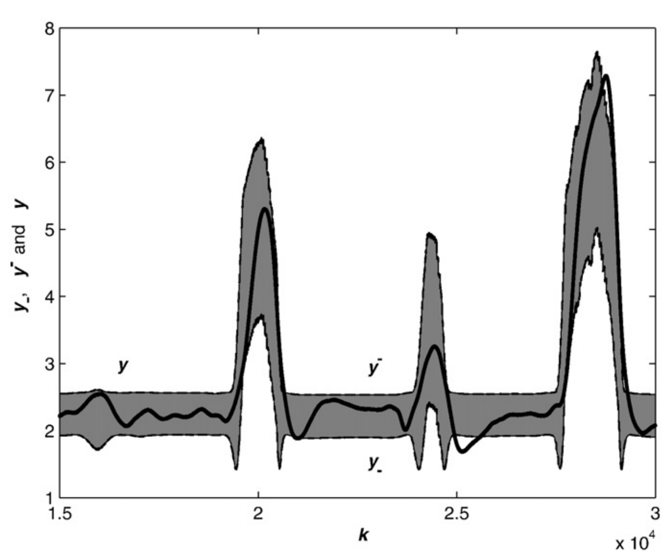

Fig. 7. The validation of fuzzy model with confidence interval.

The verification of the fuzzy model output $\hat{C}_{NH4N_{out}}$ and the process output $C_{NH4N_{out}}$ are shown in Figs. 4 and 5, where the verification of the fuzzy confidence interval is shown.

The validation of the fuzzy model and the confidence interval is shown in Figs. 6 and 7. In Fig. 6 it is evident that the error between the measured and estimated output variable is in some regions relatively large. So in those regions the model is inadequate and does not represent the behavior of the plant. The fuzzy model is, therefore, not reliable for those regions. To avoid this problem we introduced the confidence band in which we always find the output variable. The validation of this band is, with some small violations, shown in Fig. 7.

5. Conclusion

A new method of confidence-interval identification, that is applicable when a finite set of measurement data is available, has been proposed. The idea is extended to the modelling of the optimal lower and upper bound functions that define the band that contains the whole measured output variable. This results in the lower and upper fuzzy bar, which can be of great importance in the case of families of functions where the parameters of the observed system vary within certain intervals. In our example the proposed method is applied to model the confidence interval of a highly nonlinear waste-water treatment plant. The proposed approach can be used in applications such as process monitoring and fault detection, where a significant violation of the normal band means the irregular functioning of the system.

References

[1] J. Ackermann, Robust Control, Springer-Verlag, 1993.
[2] J.C. Bezdek, C. Coray, R. Gunderson, J. Watson, Detection and characterization of cluster substructure, I. Linear structure: fuzzy *c*-lines, SIAM J. Appl. Math. 40 (2) (1981) 339–357.
[3] J.C. Bezdek, C. Coray, R. Gunderson, J. Watson, Detection and characterization of cluster substructure, II. Linear structure: fuzzy *c*-varieties and convex combination thereof, SIAM J. Appl. Math. 40 (2) (1981) 358–372.
[4] J. Chiang, Fuzzy linear programming based on statistical confidence interval and interval-valued fuzzy set, Eur. J. Oper. Res. 129 (1) (2001) 65–86.
[5] R.R. Goldberg, Methods of Real Analysis, John Wiley and Sons, 1976.
[6] S. Jakubek, N. Keuth, A local neuro-fuzzy network for high-dimensional models and optimization, Eng. Appl. Artif. Intell. 19 (2005) 705–717.
[7] P. Julian, M. Jordan, A. Desages, Canonical piecewise linear approximation of smooth functions, IEEE Trans. Circuits Syst. 45 (1998) 567–571.
[8] B. Kosko, Fuzzy systems as universal approximators, IEEE Trans. Comput. 43 (11) (1994) 1329–1333.
[9] C.-H. Lin, Siso nonlinear system identification using a fuzzy-neural hybrid system, Int. J. Neural Syst. 8 (3) (1997) 325–337.
[10] I. Škrjanc, S. Blažič, O. Agamennoni, Interval fuzzy modeling applied to Wiener models with uncertainties, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 35 (5) (2005) 1092–1095.
[11] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modelling and control, IEEE Trans. Syst. Man Cybern. 15 (1985) 116–132.
[12] H.G.C. Ying, Necessary conditions for some typical fuzzy systems as universal approximators, Automatica 33 (1997) 1333–1338.
[13] L.-X. Wang, J.M. Mendel, Fuzzy basis functions, universal approximation, and orthogonal least-squares learning, IEEE Trans. Neural Netw. 3 (5) (1992) 807–814.