Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Control Engineering Practice 12 (2004) 615–624
CONTROL ENGINEERING PRACTICE
www.elsevier.com/locate/conengprac

Engine load prediction in off-road vehicles using multi-objective nonlinear identification

K. Maertens[a,b,*,1], T.A. Johansen[c], R. Babuška[a]

[a] Systems and Control Engineering Group, Department of Electrical Engineering, Delft University of Technology and Katholieke University Leuven, P.O. Box 5031, 2600 GA Delft, Netherlands

[b] Department of Agro-Engineering and Economics, Katholieke Universiteit Leuven, Kasteelpark Arenberg 30, B-3001 Leuven, Belgium

[c] Department of Engineering Cybernetics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Accepted 2 July 2003

Abstract

The multi-objective identification of nonlinear dynamic models consisting of local linear models is considered. The tradeoff between global model accuracy and local model interpretability is explicitly considered by introducing weights on the criteria for local model accuracy. A strategy is proposed to tune the local weights in order to achieve similar tradeoff for each local model. In this way, better generalization is achieved. The multi-objective identification algorithm has been applied to predict the engine load of an off-road vehicle operating under varying working load conditions. The analysis tools have proven useful for the construction of an accurate and robust engine load prediction model. The resulting model can directly be used in model-based control algorithms in automatic tuning systems that explicitly deal with constraints on the working region.

Keywords: Nonlinear system identification; Fuzzy modeling; Takagi–Sugeno model; Multi-objective optimization; Engine load prediction

1. Introduction

This articles addresses the tradeoff between the interpretability of the local models as linearizations of the nonlinear system and the prediction performance of the global nonlinear model. The objective is to identify smooth Takagi–Sugeno fuzzy models with local models that have valid interpretations while minimizing the model's global prediction performance. This problem is relevant in many nonlinear control applications (Roubos, Mollov, Babuška, & Verbruggen, 1999; Fisher & Nelles, 1998; Shorten, Murray-Smith, Bjørgan, & Gollee, 1999; Johansen, Shorten, & Murray-Smith, 2000) and in many prediction problems (Yen, Wang, & Gillespie, 1998). A multi-objective identification algorithm, minimizing a weighted sum of global and local prediction error criteria, was suggested by Yen et al. (1998) for Takagi–Sugeno fuzzy models. Tools for the analysis of conflicts among the objectives and thus methods for selecting the weights or priorities among conflicting objectives were provided in Johansen and Babuška (2002). The present work extends this method in the sense that it provides means for selecting weights and priorities to automatically balance the conflicts between objectives.

This multi-objective identification algorithm has been used to identify a dynamic, multiple input Takagi–Sugeno fuzzy model for engine load prediction of combine harvesters. These types of off-road vehicles are typically operating in a wide range of conditions in which the engine occasionally acts as a limiting factor and, therefore, the engine load should be restricted when applying automatic control schemes to tune the machine speed or other harvester settings. A prediction model makes it possible to take this constraint explicitly into account (Maciejowski, 2002). The same methodology can be used to predict other machine parameters and can be applied to other types of off-road vehicles.

*Corresponding author. Systems and Control Engineering Group, Laboratory for Agro-Machinery and Processing, Delft University of Technology and Katholieke University Leuven, Kasteelpark Arenberg 30, Leuven 3001, Belgium. Tel.: +32-163-21444; fax: +32-163-28590.
E-mail address: koen.maertens@agr.kuleuven.ac.be (K. Maertens).

[1] The author gratefully acknowledges I.W.T. for the support through doctoral Grant No. 3249 and F.W.O. for financing his stay at Delft University of Technology.

0967-0661/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0967-0661(03)00143-6

Section 2 gives an overview of the standard least-squares criteria for estimating the consequent parameters of the Takagi–Sugeno models. A multi-objective criterion is presented that makes a tradeoff between the transparency of the local models and the global accuracy of the fuzzy model. It also allows one to interpret the resulting parameter vectors in terms of conflict between local and global accuracy. Section 3 illustrates how this algorithm can be used to model engine load responses of off-road vehicles with physically interpretable local parameter vectors and without loss of global accuracy. The same strategy can directly be applied to other nonlinear dynamic systems. Section 4 contains a discussion of the results and conclusions are drawn in Section 5.

2. Methods

2.1. Takagi–Sugeno fuzzy model

The identification of dynamic Takagi–Sugeno fuzzy input/output models of the following form is considered:

$$y(k) = \sum_{i=1}^{N} [(1 - A_i(q^{-1}))y(k) + B_i(q^{-1})u(k)) + d_i]w_i(z(k)) + e(k), \quad (1)$$

where $u(k) \in R^r$ denotes the input, $y(k) \in R^m$ is the output, $e(k) \in R^m$ accounts for unmodeled phenomena, N defines the number of rules/local models and functions $w_i : R^p \to [0, 1]$ define the degree of fulfilment for a given set of premise variables $z(k) \in R^p$. The individual models are determined by the polynomials $A_i(q^{-1}) = 1 + a_{1,i}q^{-1} + \cdots + a_{n_y,i}q^{-n_y}$ and $B_i(q^{-1}) = b_{0,i} + \cdots + b_{n_u,i}q^{-n_u}$. The vector $z(k)$, which contains the premise variables, is a subset of the information vector

$$\tilde{\psi}(k) = (-y(k-1), \ldots, -y(k-n_y), u(k), \ldots, u(k-n_u))^{\mathrm{T}}. \quad (2)$$

Fuzzy models of form (1) result from fuzzy inference on a set of fuzzy rules

$$\text{IF} \quad z(k) \in Z_i \quad \text{THEN} \quad y(k) = (1 - A_i(q^{-1}))y(k) + B_i(q^{-1})u(k) + d_i, \quad (3)$$

where the premise is defined by a fuzzy set Z_i defined by the membership functions $\mu_i : R^p \to [0, 1]$ and the consequent is a local linear dynamic model. The function w_i is given by

$$w_i(z) = \frac{\mu_i(z)}{\sum_{j=1}^{N} \mu_j(z)}, \quad (4)$$

thus fulfilling the condition

$$\sum_{i=1}^{N} w_i(z) = 1, \quad (5)$$

see Takagi and Sugeno (1985) for details. The only assumption we make on the set of fuzzy rules is that it is complete in the sense that $\mu_j(z) > 0$ for some j for all z, such that (4) is well defined. Eq. (1) can be reformulated into a form that is more convenient for system identification by introducing the definitions

$$\psi(k) = \begin{pmatrix} \tilde{\psi}(k) \\ 1 \end{pmatrix}, \quad (6)$$

$$\theta_i = (a_{1,i}, \ldots, a_{n_y,i}, b_{0,i}, \ldots, b_{n_u,i}, d_i)^{\mathrm{T}}, \quad (7)$$

where $\psi(k)$ is the information vector (2) augmented with a constant element, and θ_i are the possibly unknown parameters associated with the local linear model of the ith rule. As a consequence, the general input/output relation for the full model can be written as function of the local model parameters

$$y(k) = \sum_{i=1}^{N} \psi^{\mathrm{T}}(k)\theta_i w_i(z(k)) + e(k). \quad (8)$$

Furthermore, defining

$$\varphi(k) = \begin{pmatrix} \psi(k)w_1(z(k)) \\ \vdots \\ \psi(k)w_N(z(k)) \end{pmatrix}, \qquad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix}, \quad (9)$$

the linear regression can be written as

$$y(k) = \varphi^{\mathrm{T}}(k)\theta + e(k). \quad (10)$$

For the purpose of system identification, assume that an input–output data set with n samples is available. The fuzzy partition of the antecedent space, which is defined by the membership functions μ_i, is obtained in a previous identification step. An overview of possible strategies to construct this antecedent partition is given in Babuška (1998).

In the remainder of this section, we briefly review the well-known global and locally weighted identification methods to calculate consequent parameter vector θ and present an alternative algorithm to make an explicit tradeoff between global and local optimization.

2.2. Global identification criterion

The objective of this algorithm is to identify the local model parameters $\theta_1, \ldots, \theta_N$ that give a global model with the best prediction performance Takagi and Sugeno (1985):

$$V(\theta) = \frac{1}{n} \sum_{k=1}^{n} (y(k) - \varphi^{\mathrm{T}}(k)\theta)^2. \quad (11)$$

This problem can be solved by using a standard least-squares approach.

2.3. Locally weighted identification criterion

The objective of this algorithm is to identify the local model parameters $\theta_1, \ldots, \theta_N$ that give local models which are close local approximations to the underlying nonlinear system (Johansen & Foss, 1993). A weighted least-squares prediction error criterion is associated with each local model

$$V_i(\theta_i) = \frac{1}{n}\sum_{k=1}^{n} (y(k) - \psi^{\mathrm{T}}(k)\theta_i)^2 w_i(z(k)). \tag{12}$$

The weighting factor $w_i(z(k))$ ensures that the parameters θ_i are influenced only by the data points within the fuzzy set Z_i that defines the region of validity of the ith local model.

2.4. Multi-objective identification criterion

2.4.1. Algorithm

It was suggested in Yen et al. (1998) to minimize the weighted sum of the global and local identification criteria (11) and (12).

$$\min_{\theta}\left(V(\theta) + \sum_{i=1}^{N} \beta_i V_i(\theta_i)\right). \tag{13}$$

The solution can be found by using a least-squares approach. The weighting parameters $\beta_i \geqslant 0$ parameterize the set of Pareto-optimal solutions of the underlying multi-objective optimization problem, and essentially determine the tradeoff between the possibly conflicting objectives of global model accuracy and local model interpretability.

2.4.2. Conflict analysis

The selection $\beta_i = 1\ \forall i$ in multi-objective criterion (13) will in general give a fairly balanced tradeoff. It is still of interest to study in detail how the choice of β influences the tradeoff. In particular, it is of interest to analyze the degree of conflict between the different objectives in (13) for a given data sequence and different values of β. This will provide the user with information that can be used to validate the model and data, as well as to improve the model, not only by selecting a better values of β_i, but possibly also by modifying the model structure or membership functions, or by adding/removing constraints.

Let the minimum of (13) be denoted $\hat{\theta}(\beta)$ for a given data sequence and a given vector of weights β. The minimum of (13) satisfies the Karush–Kuhn–Tucker (KKT) conditions

$$\frac{\partial V}{\partial \theta}(\hat{\theta}(\beta)) + \sum_{i=1}^{N} \beta_i \frac{\partial V_i}{\partial \theta}(\hat{\theta}(\beta)) = 0. \tag{14}$$

If there are no conflicts among the objectives and constraints, i.e., $\hat{\theta}(\beta)$ minimizes all of the individual objectives simultaneously then each of the terms in (14) will be zero. If there are conflicts, however, the directions and lengths of each of the (vector) terms of (14) will indicate the degree of conflict and which objectives are actually in conflict with each other.

Considering the parameter vector θ_i of the ith local model, Eq. (14) leads to

$$\frac{\partial V}{\partial \theta_i}(\hat{\theta}(\beta)) + \beta_i \frac{\partial V_i}{\partial \theta_i}(\hat{\theta}(\beta)) = 0. \tag{15}$$

Next, define the following sensitivity measures associated with each parameter $\theta_{i,j}$, which is the jth parameter in the ith local model:

$$\pi^g_{i,j}(\beta) = -\frac{\partial V}{\partial \theta_{i,j}}(\hat{\theta}(\beta)), \tag{16}$$

$$\pi^l_{i,j}(\beta) = -\frac{\partial V_i}{\partial \theta_{i,j}}(\hat{\theta}(\beta)) = \frac{1}{\beta_i}\frac{\partial V}{\partial \theta_{i,j}}(\hat{\theta}(\beta)), \tag{17}$$

which can easily be computed from the regressor matrices. The quantity $\pi^g_{i,j}(\beta)$ can be interpreted as the decrease in the global identification criterion V that can be achieved by a small increase in $\hat{\theta}_{i,j}(\beta)$. Likewise, the quantity $\pi^l_{i,j}(\beta)$ can be interpreted as the decrease in the local identification criterion V_i that can be achieved by a small increase in $\hat{\theta}_{i,j}(\beta)$. Hence, large values of $\pi^g_{i,j}(\beta)$ and $\pi^l_{i,j}(\beta)$ indicate conflicts between global and local performance.

Notice that due to (15)

$$\pi^g_{i,j}(\beta) + \beta_i \pi^l_{i,j}(\beta) = 0. \tag{18}$$

A larger local model weight β_i will bring about a smaller local sensitivity parameter $\pi^l_{i,j}(\beta)$, but will result in a higher global sensitivity $\pi^g_{i,j}(\beta)$. The influence of β_i on both sensitivity measures of one particular consequent parameter is illustrated in Fig. 1. This consequent parameter is quasi-optimal in global least-squares sense as defined by Eq. (11) when a β_i value of less than 10^{-2} is chosen, while a β_i value of more than 10 approximates the local least-squares solution of Eq. (12).

The analysis of $\pi^g_{i,j}(\beta)$ and $\pi^l_{i,j}(\beta)$ can provide the user with significant information about the model and data, as illustrated in Section 3.4.

2.4.3. Choice of weighting parameters β_i

The β_i parameters, which define the weights on the local model accuracy in criterion (13), have to be determined before the sensitivity measures $\pi^g_{i,j}(\beta)$ and

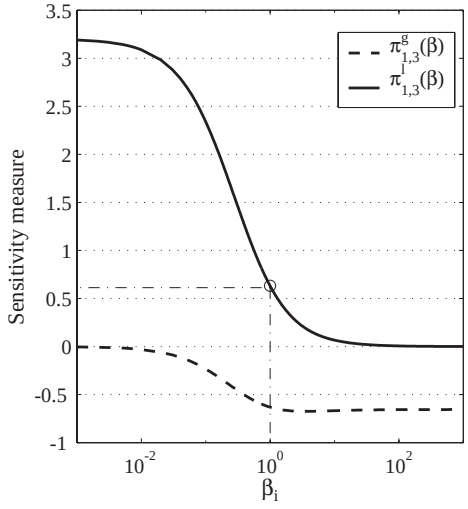

Fig. 1. Influence of the weight β_i on the local and global sensitivity measures of parameter $a_{1,3}$, related to the multi-objective consequent estimation of Tables 2, 3 and 4. The marker corresponds to the local sensitivity parameter $\pi^l_{1,3}$ of Fig. 5.

$\pi^l_{i,j}(\beta)$ can be calculated. Different strategies are possible to design the weighting vector β:

- The same value is chosen for each β_i, resulting in a consequent parameter set that makes a global trade-off between local and global accuracy. The resulting sensitivity measures are directly related to the partition of the antecedent space.
- Each individual β_i parameter is determined according to the local need for accuracy. The sensitivity parameters are determined by both the membership functions and the distribution of the local model weights.
- The following quadratic measure Π^l_i can been used to quantify the local sensitivity of one particular local model:

$$\Pi^l_i = \sqrt{\frac{1}{n_{\theta_i}} \sum_{j=1}^{n_{\theta_i}} (\pi^l_{i,j})^2}, \tag{19}$$

where n_{θ_i} denotes the number of parameters of the ith local model. The set of N model sensitivity measures Π^l_i can be balanced by tuning the N model weights β_i iteratively under the constraint of a fixed mean local weight value β^*,

$$\frac{1}{N} \sum_{i=1}^{N} \beta_i \equiv \beta^*. \tag{20}$$

This average model weight β^* determines the tradeoff between model transparency and accuracy of the total Takagi–Sugeno model. The optimized vector β reveals the local conflict between transparency and global accuracy and is determined by the partition of the antecedent space. The balanced Π^l_i measures guarantee physically interpretable consequent parameters θ_i with sufficient global accuracy. As a consequence, a better generalization of the global system can be expected, making the fuzzy model more reliable for new data that have not been used for training.

The appropriate choice of the weights β_i depends on the objective of the fuzzy model, the specific application and the available process knowledge. A constant β_i vector can be used to evaluate the antecedent partition, a user-defined local weighting is of particular interest in case of good a priori process knowledge and a specific problem definition, while the strategy of balanced local sensitivity parameters can be used to generate a robust and accurate model performance across the entire premise space without the need for a priori process knowledge.

3. Engine load prediction for off-road vehicles

Engines of large off-road vehicles such as dozers, harvesters or mobile equipment for road construction, have to provide power for moving this heavy equipment, but are also loaded with some additional, time-varying tasks that depend on the particular type of equipment. A combine harvester, for example, has to move through the field, but has also the task of cutting, processing and storing the crop. The latter set of tasks can consume more than 70% of the total engine power. As a consequence, the engine load can approach its maximum allowable limit in some operating regimes and, therefore, precautions must be taken to avoid engine overload, especially under automatic ground speed control of the vehicle. An accurate model to predict the engine load as a function of the machine load signals and the ground speed input is essential to design a constrained model-based control algorithm (Maciejowski, 2002). Commercial combine harvesters are commonly equipped with a controller area network (CAN) and an embedded software platform that makes the integration of control algorithms easy and reliable.

The multi-objective identification criterion of Section 2.4 will be applied to predict the engine load of a combine harvester on the basis of the feedrate, machine slope and the ground speed input signals. A tradeoff will be made between global accuracy and local transparency to guarantee robust prediction performance. The local linear representations of the global nonlinear system will make it possible to analyze the control problem locally with standard tools for linear systems and in addition, the interpretable local models can directly be implemented into an expert system for fault detection purposes.

Table 1
Selected model inputs for engine load prediction of the combine harvester

Signal	Symbol	Units
Engine load	$y(k)$	%
Feedrate	$u_F(k-4)$	V
Slope	$u_S(k-1)$	deg
Handle position	$u_H(k-1)$	%

3.1. Model structure

Table 1 gives an overview of the signals that are closely related with the engine load of the harvester:

- The feedrate signal $u_F(k-4)$ is obtained by registering the torque on the driving belt of the combine header that cuts and transports the crop toward the internal part of the harvester. This signal is an indication of the crop throughput, about 0.6 s before this biomass is actually processed by the internal separation elements (Maertens, De Baerdemaeker, Ramon, & De Keyser, 2001).
- The backward/forward machine slope $u_S(k-1)$ immediately determines the engine load via the weight of the harvester and the collected crop.
- The ground speed of mobile agricultural machinery is manually varied by means of a handle. The position of this handle $u_H(k-1)$ is the most important input variable to predict the engine load and is also the main control parameter that can be varied by automatic grain loss control systems.
- The engine load $y(k)$ itself is estimated indirectly by measuring the instantaneous fuel consumption of the diesel engine and is characterized by high-frequency noise terms.

These four input signals have been selected from a larger set of measured machine parameters on the basis of physical insight and by applying an iterative selection scheme that makes use of a genetic algorithm to calculate the significance of each input signal.

All these measurements were carried out via the standard, commercial CAN at a sample rate of 5 Hz. Appropriate hardware filters are installed to prevent aliasing.

Based on physical process knowledge and results from correlation analysis, the following dynamic multiple input single output (MISO) structure has been chosen for the fuzzy rules

$$\text{IF} \quad (u_F(k-4)\in Z_{F,i}) \quad \text{AND} \quad (u_S(k-1)\in Z_{S,i}),$$
$$\text{THEN}$$
$$y(k) = a_{1,i}y(k-1) + a_{2,i}y(k-2) + b_{F,i}u_F(k-4) + b_{S,i}u_S(k-1) + b_{H,i}u_H(k-1) + d_i. \quad (21)$$

The premise variables are the feedrate $u_F(k-4)$ and the machine slope $u_S(k-1)$, while two lagged outputs and the handle position $u_H(k-1)$ are also added to the inputs of the consequent rules. The AND operation has been realized as the product of both membership degrees. This structure has been chosen on the basis of physical insights and some iterative manual tuning. For instance, the two lagged outputs $y(k-1)$ and $y(k-2)$ are necessary to capture the system's dynamics, but appear to be redundant with respect to the partition of the premise space.

3.2. Construction of membership functions

A training set of 5185 data samples has been used to partition the antecedent feedrate-slope space into regimes that provide a locally linear engine load response. The Gustafson–Kessel fuzzy clustering algorithm (Gustafson & Kessel, 1979) has been applied for different numbers of clusters according to the methodology described in Babuška (1998). As a result, three clusters were found which after projection provide the partition of the antecedent space. They are depicted in Fig. 2 together with the cluster centers and the projected samples of the training data set. These clusters were found in an automatic manner, but at the same time, a good a posteriori interpretation can be given to the obtained partition. The first cluster characterizes a low feedrate and corresponds to the machine regime at which the vehicle is only moving through the field without actually harvesting any crop. The second and third cluster correspond to engine load responses registered for downhill and uphill harvesting, respectively.

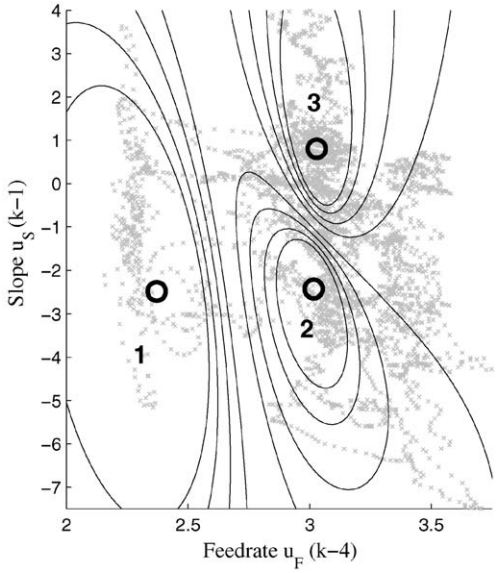

Fig. 2. The data set projected on the antecedent space together with resulting clusters. The contour lines correspond to membership degrees of 1, 0.95, 0.9, 0.8 and 0.65, respectively.

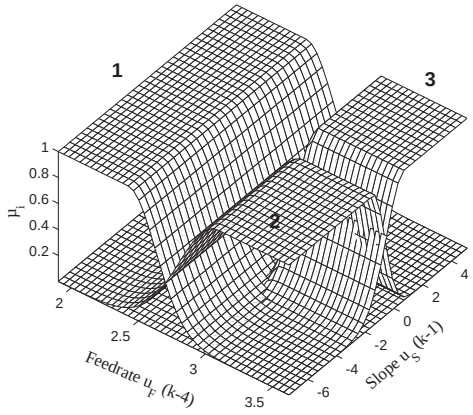

Fig. 3. The resulting fuzzy partition of the antecedent space.

Fig. 3 shows the Cartesian product space intersection of the membership functions that approximate the fuzzy partition of Fig. 2. Piecewise exponential membership functions with the following structure are used:

$$\mu(z;c_l,c_r,w_l,w_r)=\begin{cases}\exp\left(-\left(\dfrac{z-c_l}{2w_l}\right)\right) & \text{if } z<c_l,\\ \exp\left(-\left(\dfrac{z-c_r}{2w_r}\right)\right) & \text{if } z>c_r,\\ 1 & \text{otherwise,}\end{cases} \quad (22)$$

where c_l and c_r are the left and right shoulder, respectively, and w_l, w_r are the left and right width.

The membership functions were manually extended toward the boundaries of the premise space in order to guarantee a sufficient coverage for the application of the model in new situations that are not fully covered by the training set of Fig. 2.

3.3. Multi-objective consequent parameter estimation

Once the membership functions $\mu_i(z)$ are constructed, local weights $w_i(z)$ can be calculated by Eq. (4) and the consequent parameter vectors θ_i can be estimated by choosing one of the criteria of Section 2.

The consequent parameters found by solving the quadratic program resulting from criterion (13) are given in Tables 2–4 for a constant set of local weights of $\beta_i = 10^{-3}$, 1 and 10^3, respectively.

An overview of the prediction accuracy for the different choices of β_i is given in Table 5 for both the training data set and a different validation set. The root mean squared error (*RMSE*) has been used to evaluate the prediction accuracy of the different parameter sets

$$RMSE=\sqrt{\frac{1}{n}\sum_{k=1}^{n}(y(k)-y_p(k))^2}, \quad (23)$$

where $y_p(k)$ and $y(k)$ denote the predicted and measured output, respectively.

Table 2
Consequent parameters obtained by multi-objective identification with main emphasis on global least-squares accuracy ($\beta_i = 10^{-3}$, validation set $RMSE = 4.805\%$)

i	$a_{1,i}$	$a_{2,i}$	$b_{F,i}$	$b_{S,i}$	$b_{H,i}$	d_i
1	1.57	−0.686	2.86	−0.011	0.070	−5.05
2	0.93	−0.203	8.50	0.853	0.128	−15.3
3	0.93	−0.164	8.18	0.946	0.237	−23.7

Table 3
Consequent parameters obtained by weighted multi-objective identification ($\beta_i = 1$, validation set $RMSE = 4.802\%$)

i	$a_{1,i}$	$a_{2,i}$	$b_{F,i}$	$b_{S,i}$	$b_{H,i}$	d_i
1	1.53	−0.640	1.57	0.057	0.066	−1.75
2	0.94	−0.201	7.47	0.733	0.122	−13.1
3	0.94	−0.157	7.62	0.810	0.210	−20.8

Table 4
Consequent parameters obtained by multi-objective identification with emphasis on local least-squares accuracy ($\beta_i = 10^3$, validation set $RMSE = 4.833\%$)

i	$a_{1,i}$	$a_{2,i}$	$b_{F,i}$	$b_{S,i}$	$b_{H,i}$	d_i
1	1.42	−0.540	1.31	0.082	0.068	−1.09
2	0.95	−0.197	6.94	0.678	0.122	−12.1
3	0.94	−0.154	7.16	0.767	0.191	−18.5

Table 5
Prediction accuracy of different Takagi–Sugeno models for training and validation data, for both the initial and ANFIS optimized antecedent partitions

	Unbalanced $\pi^l_{j,i}$		Balanced $\pi^l_{j,i}$	
	Training (%)	Validation (%)	Training (%)	Validation (%)
Linear	3.489	5.509	3.489	5.509
$\beta^a = 10^{-3}$	2.990	4.805	2.981	4.728
$\beta^a = 10^{-1}$	2.993	4.789	2.993	4.715
$\beta^a = 1$	3.038	4.802	3.032	4.727
$\beta^a = 10^1$	3.075	4.828	3.062	4.749
$\beta^a = 10^3$	3.083	4.833	3.068	4.755
ANFIS[a]	2.923	4.864	2.923	4.864
$\beta^a = 10^{-3}$	2.924	4.855	2.925	4.845
$\beta^a = 10^{-1}$	2.949	4.727	2.952	4.727
$\beta^a = 1$	3.015	4.737	2.983	4.701
$\beta^a = 10^1$	3.085	4.794	3.089	4.794
$\beta^a = 10^3$	3.105	4.806	3.105	4.806

[a] Standard ANFIS algorithm according to Jang (1993).

The different sets of consequent parameters and the resulting model accuracies illustrate the tradeoff between model transparency and global accuracy:

- The consequent parameters of Table 2 result from multi-objective optimization with only a small emphasis on the local model performance

$V_i(\theta_i)$ ($\beta_i = 10^{-3}$). The static gain parameter $b_{S,1}$ (-0.011) is clearly not physically interpretable, since an increasing machine slope results in a higher engine load (a positive parameter value). The same set of consequent parameters $\hat{\theta}(\beta)$ provides the lowest training error. This is a typical example of the conflict between global accuracy and local interpretability in Takagi–Sugeno models (see also the validation results in Table 5).

- The opposite holds for the parameter set of Table 4, which corresponds to $\beta_i = 10^3$. The resulting parameters are physically interpretable, but a relatively large training error is obtained.

Although the relation between the local model weights β_i and the accompanying training error is monotonously increasing for the training data, a minimum is found when applying the same model to a validation set. This minimum illustrates the usefulness of the general tradeoff between global and local accuracy to achieve a better generalization of the system description. The consistently lower validation error after balancing the local sensitivity parameters reveals the importance of tuning the individual local model weights β_i to achieve a balanced tradeoff for every local model i (Table 5).

This tradeoff between local transparency and global accuracy as a function of the weights β_i, is mainly determined by the partition of the antecedent space. To show the influence of a different partition, the adaptive-network-based fuzzy inference system (ANFIS) method (Jang, 1993) has been used to optimize the initial membership functions of Fig. 3.

The resulting membership functions of the ANFIS algorithm are shown in Fig. 4 and the corresponding prediction accuracies for the different β_i values are included in Table 5. Again, an optimal β_i value is found that makes a tradeoff between training accuracy and generalization. The consequent parameters for the standard ANFIS configuration ($\beta_i = +\infty$) are presented in Table 6 and the prediction accuracy is also provided in Table 5.

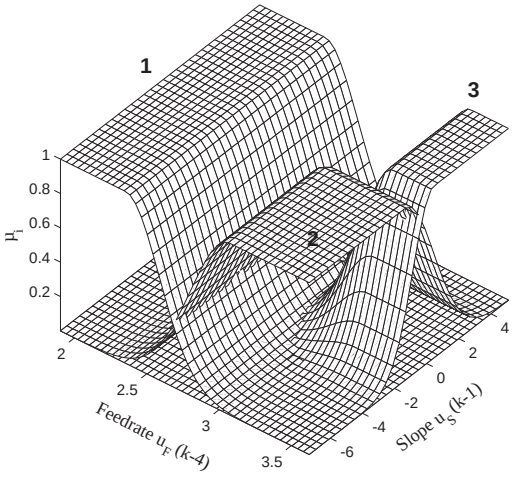

Fig. 4. Membership functions obtained by ANFIS optimization (using the partition of Fig. 3 as the initial partition).

3.4. Conflict analysis

The consequent parameters that are globally ($\beta_i = +\infty$) optimized with the standard ANFIS method have shown a relatively poor generalization performance when applied to a different set, illustrated by a relatively large difference between training and validation error. However, no strong contradictions can be found between the consequent parameter values and physical process knowledge. Therefore, an objective quantitative measure that indicates strong conflicts between global and local model performance without the need for an additional data set is of particular interest during the design of the model structure and the evaluation of the consequent parameters.

Fig. 5 shows the local sensitivity measures $\pi^l_{i,j}$ (17) of the initial partition, related to the estimated consequent

Table 6
Consequent parameters obtained by the global ($\beta_i = +\infty$) ANFIS optimization algorithm corresponding to the ANFIS optimized membership functions of Fig. 4 (Validation set $RMSE = 4.864\%$)

i	$a_{1,i}$	$a_{2,i}$	$b_{F,i}$	$b_{S,i}$	$b_{H,i}$	d_i
1	1.57	−0.679	3.61	0.011	0.068	−6.89
2	0.91	−0.185	9.88	0.899	0.111	−18.1
3	0.89	−0.139	12.9	1.37	0.464	−51.1

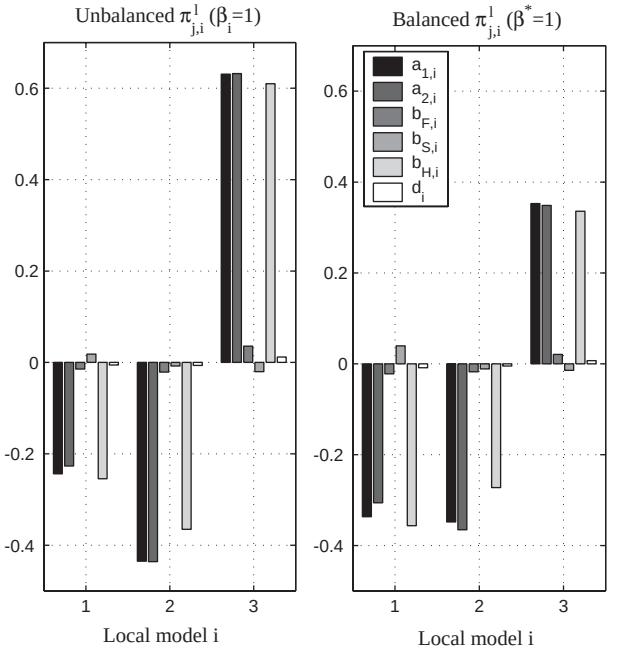

Fig. 5. Local criterion sensitivity parameters $\pi^l_{i,j}$ for the initial membership functions of Fig. 3 before ($\beta_i = 1$) and after ($\beta^* = 1$) local conflict balancing.

parameters of Table 3 ($\beta_i = 1$). The conflicts between local and global accuracy are mainly present at consequent parameters $a_{1,i}$ and $a_{2,i}$ of the lagged output sequence $y(k)$ and the static gain parameter $b_{H,i}$ of the handle input $u_H(k-1)$.

The local sensitivity measures for the ANFIS optimized antecedent partition of Fig. 4 are presented in the left part of Fig. 7. Again, important local sensitivity measures are found for the consequent parameters $a_{1,i}$, $a_{2,i}$ and $b_{H,i}$.

3.5. Balanced local sensitivity measures

The previous section has illustrated the role of the local sensitivity parameters $\pi^l_{i,j}$ as indicators for conflicts between local and global accuracy. Important conflicts between transparency and accuracy indicate poor generalization and, consequently, may result in poor prediction performance for new data.

The right part of Fig. 5 shows the result after balancing the local sensitivity measure Π^l_i (19) of the antecedent partition of Fig. 3. The accompanying consequent parameters given in Table 7 were estimated with $\beta = (0.29, 0.74, 1.97)^{\mathrm{T}}$. This new weight vector has lowered the highest $\pi^l_{j,i}$ values by putting a higher weight β_3 on the local accuracy of the third model. The main conflicts remain still at the consequent parameters $a_{1,i}$, $a_{2,i}$ and $b_{H,i}$, which characterize the engine load response to $\Delta u_H(k-1)$ variations. The step responses of the three models before and after balancing the local sensitivity parameters are plotted in Fig. 6. The largest difference is visible in the static gain of model 2, calculated as $b_{H,2}/(1 - a_{1,2} - a_{2,2})$. The model accuracies after balancing and for different β^* choices are compared in Table 5.

The same balancing algorithm has been applied for the ANFIS optimized membership functions of Fig. 4 and the sensitivity parameters are compared in Fig. 7. A β vector of $(1.48, 1.31, 0.28)^{\mathrm{T}}$ has been obtained to achieve the balanced local conflicts and the corresponding consequent parameters are shown in Table 8. The prediction results of the balanced models are included in Table 5.

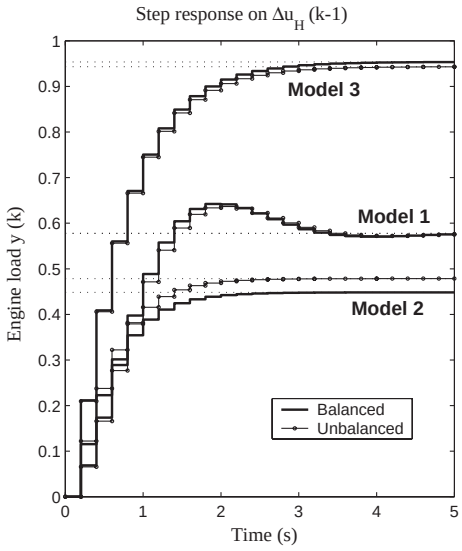

Fig. 6. Step responses of three local models based on the initial partition of Fig. 3.

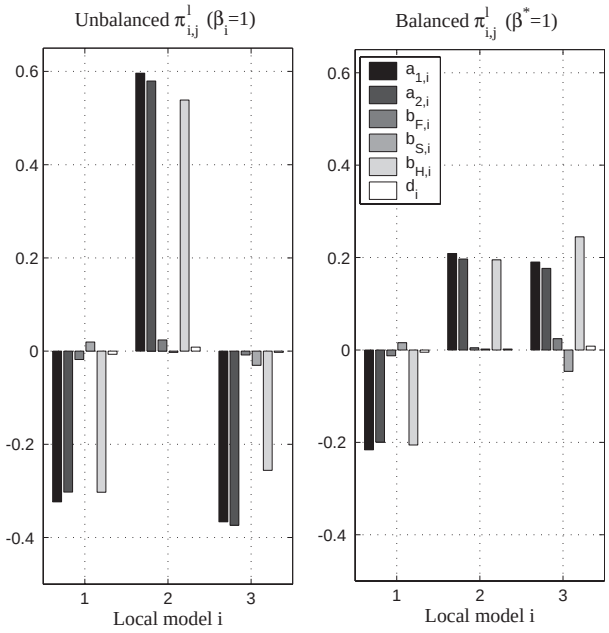

Fig. 7. Local criterion sensitivity parameters $\pi^l_{i,j}$ for the ANFIS-optimized antecedent partition of Fig. 4 before ($\beta_i = 1$) and after ($\beta^* = 1$) applying the balancing algorithm.

Table 7
Consequent parameters obtained by multi-objective optimization that correspond to the antecedent partition of Fig. 3 and result from the balanced local sensitivity parameters shown in the right part of Fig. 5 (Validation set $RMSE = 4.727\%$)

i	$a_{1,i}$	$a_{2,i}$	$b_{F,i}$	$b_{S,i}$	$b_{H,i}$	d_i
1	1.52	−0.642	2.14	0.062	0.069	−3.05
2	0.93	−0.191	7.80	0.750	0.115	−13.6
3	0.93	−0.155	7.61	0.792	0.211	−20.9

Table 8
Consequent parameters for the ANFIS-optimized membership functions of Fig. 4, resulting from the balanced sensitivity parameters shown in the right part of Fig. 7, under the constraint $\beta^* = 1$ (validation set $RMSE = 4.701\%$)

i	$a_{1,i}$	$a_{2,i}$	$b_{F,i}$	$b_{S,i}$	$b_{H,i}$	d_i
1	1.51	−0.625	2.07	0.109	0.065	−2.84
2	0.94	−0.186	7.36	0.676	0.125	−13.5
3	0.92	−0.139	7.79	0.814	0.264	−24.7

4. Discussion of results

4.1. Local sensitivity measures

Two different strategies for designing the weight vector β have been applied to the design of the engine load prediction model.

4.1.1. Equally distributed β_i weights

The local accuracy of all local models is weighted with the same β_i value and the resulting sensitivity parameters are used to evaluate the partition of the antecedent space as illustrated in Fig. 5 for the original membership functions and in Fig. 7 for the ANFIS-optimized partition. No particular local conflicts are found for any of the local models and both partitions are expected to give good prediction results with interpretable consequent parameters.

4.1.2. Balanced local sensitivity measures

The local sensitivity measures Π_i^l have been balanced for both antecedent partition of Figs. 3 and 4. The validation of the resulting models on a different data set has illustrated the improved generalization for both partitions (Table 5). A consistently lower validation *RMSE* is found for every choice of the mean local weight β^* and for both antecedent partitions.

4.2. Model interpretation

The membership functions of both the original partition of Fig. 3 and the ANFIS optimized partition of Fig. 4 are revealing a similar regime decomposition. The physical interpretation of these three models makes it possible to analyze the corresponding consequent models that are presented by their step responses in Fig. 6. The numerical values of the static gains are given in Table 9.

Local model 1: It approximates the engine load responses when the machine is not harvesting any crop or only small amounts. The recorded feedrate variations are mainly induced by noise disturbances and unwanted correlation with other varying signals (e.g., the machine slope) and are therefore less reliable for predicting the engine load. The engine is only slightly loaded in this regime, which results in faster and less damped dynamics.

Local model 2: It characterizes the system response during downhill harvesting. The extra harvesting task introduces a higher engine load and, therefore, slower dynamics. The negative machine slope makes it easier for the engine to respond to variations of the ground speed input signal $u_H(k-1)$.

Local model 3: It describes the highest engine load signal because of the high feedrate values and positive machine slopes. This combination makes the engine load more sensitive for handle variations. The dynamics of the engine load response are similar to those of local model 2.

Table 9
Static gains of the different input variables, corresponding to the balanced consequent parameters of Table 7

i	K_F^{stat} (%/V)	K_S^{stat} (%/deg)	K_H^{stat}
1	17.54	0.50	0.57
2	29.89	2.87	0.44
3	33.82	3.52	0.94

4.3. Prediction accuracy

Fig. 8 shows the prediction performance of the balanced consequent parameters of Table 8 on a data

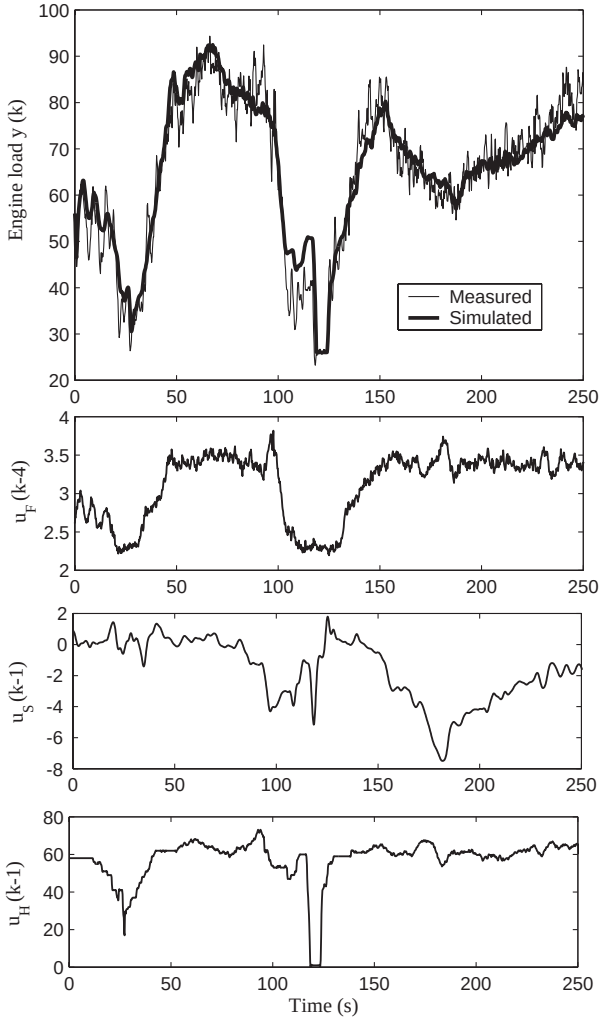

Fig. 8. Engine load prediction for a validation data set based on the ANFIS optimized antecedent space of Fig. 4 and the balanced ($\beta^* = 1$) consequent parameters of Table 8 (validation set $RMSE = 4.701\%$).

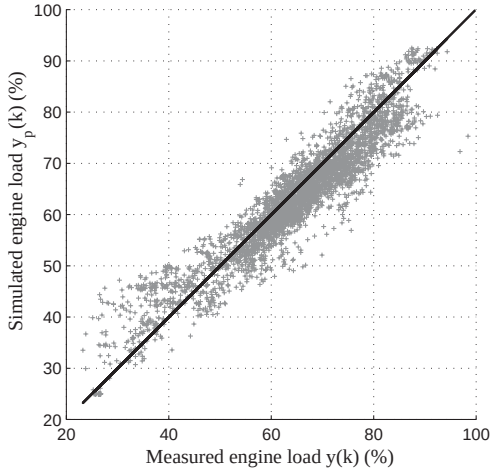

Fig. 9. A comparison of the measured and simulated engine load for the validation data set, making use of the ANFIS optimized membership functions and balanced model weights ($\beta^* = 1$, validation set $RMSE = 4.701\%$).

set that is different from the training set. The Takagi–Sugeno fuzzy model is based on the ANFIS optimized membership functions of Fig. 4 and are balanced for an average model weight $\beta^* = 1$.

The high-frequency noise distortions on the engine load measurement are obvious and are responsible for a significant part of the residual terms of the calculated training and validation *RMSE* values of Table 5. The influence of the measurement noise is also visible in Fig. 9, visualizing the simulated versus the measured engine load values of the same validation set of Fig. 8. The largest uncertainty occurs for low engine load values and is caused by variations of other machine parameters that are not included in the input space of the model. The nonlinear characteristics, indicated by the low performance of one global linear model (Table 5) and the step responses of Fig. 6, are captured by the fuzzy model since no bias terms arise on the prediction error in the validated engine load range (25–90%).

The achieved prediction accuracy in combination with the interpretability of the local model parameters are promising for the purpose of model-based constraint handling, although closed-loop experiments are necessary to confirm the results.

5. Conclusions

A multi-objective optimization formulation of the identification problem for local model networks arises naturally due to the two conflicting objectives between local and global performance. In this paper, the multi-objective formulation proposed in Johansen and Babuška (2002) is extended by introducing a methodology to select the local model weights in order to balance the local sensitivity measures. As a result, the same local–global tradeoff is achieved for every local model, implying a better generalization performance of the final Takagi–Sugeno fuzzy model without losing global accuracy. Better prediction performance, in combination with a better physical interpretability of the local model parameters, makes the resulting fuzzy model promising as the system description in model-based control schemes.

This multi-objective identification algorithm has been applied to predict the engine load responses of an off-road vehicle based on the past values of an indirect machine load measurement, the backward/forward machine slope and the ground speed input measurement. The final Takagi–Sugeno fuzzy model has a good prediction performance with an interpretable antecedent partition and physically relevant consequent parameters. This mathematical representation can directly be implemented in model-based control algorithms, to attain a higher closed-loop performance.

References

Babuška, R. (1998). *Fuzzy modeling for control*. Boston: Kluwer Academic Publishers.

Fisher, M., & Nelles, O. (1998). Predictive control based on local linear fuzzy models. *International Journal on Systems Science*, *29*, 679–697.

Gustafson, D., & Kessel, W. (1979). Fuzzy clustering with a fuzzy covariance matrix. In *36th IEEE conference on decision and control* (*CDC 79*), San Diego, USA (pp. 665–685).

Jang, J.-S. R. (1993). Anfis: Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*, 665–685.

Johansen, T. A., & Babuška, R. (2002). On multi-objective identification of Takagi–Sugeno fuzzy model parameters. In *IFAC world congress*, Barcelona.

Johansen, T. A., & Foss, B. A. (1993). Constructing NARMAX models using ARMAX models. *International Journal on Control*, *58*, 1125–1153.

Johansen, T. A., Shorten, R., & Murray-Smith, R. (2000). On the interpretation and identification of Takagi–Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, *8*, 297–313.

Maciejowski, J. M. (2002). *Predictive control with constraints*. Upper Saddle River, NJ: Prentice-Hall.

Maertens, K., Baerdemaeker, J. De, Ramon, H., & De Keyser, R. (2001). An analytical grain flow model of a combine harvester. Part 1: Design of the model. *Journal of Agricultural Engineering Research*, *79*, 55–63.

Roubos, H., Mollov, S., Babuška, R., & Verbruggen, H. B. (1999). Fuzzy model based predictive control by using Takagi–Sugeno fuzzy models. *International Journal on Approximate Reasoning*, *22*, 3–30.

Shorten, R., Murray-Smith, R., Bjørgan, R., & Gollee, H. (1999). On the interpretation of local models in blended multiple model structures. *International Journal on Control*, *72*, 620–628.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, *15*, 116–132.

Yen, J., Wang, L., & Gillespie, C. W. (1998). Improving the interpretability of TSK fuzzy models by combining global learning and local learning. *IEEE Transactions on Fuzzy Systems*, *6*, 530–537.