Evolving Fuzzy-Model-Based Design of Experiments With Supervised Hierarchical Clustering

Igor Škrjanc

Abstract—**This paper presents a new approach to design of experiments (DoE), based on an evolving fuzzy model structure and a supervised hierarchical clustering algorithm. DoE is the field that deals with the problem of how to design the most optimal and economic experimentation. The goal is to identify a highly nonlinear and possibly high-dimensional system, together with the minimal experimental effort required. The theory is well developed for linear and polynomial models; however, they are often not suitable for general use. For this reason, a fuzzy model in the form of Takagi–Sugeno (T–S) is used, because it has the properties of a universal approximator. The method works iteratively by sampling the system in the input domain and evolving the fuzzy model. The method is demonstrated with a simulation, which shows the potential of the proposed approach.**

Index Terms—**Design of experiments (DoE), fuzzy clustering, fuzzy model identification.**

I. INTRODUCTION

SYSTEM modeling and experimentation are essential parts of the engineering discipline. Such models are used for the analysis of systems and for behavior prediction under different input conditions. The classical approach to modeling is the use of basic laws to determine the model of the system. The model is presented as a mathematical function of the input variables and parameters. In order to determine the parameters of the model, we usually have to conduct experiments. In practice, a system often cannot be modeled in the classical way using theoretical knowledge. The identification of such systems is usually data-driven. Models that are often used, especially when the system is nonlinear, are neural networks, local model networks, and fuzzy models [1], [2].

If a model of the system is known, we can determine the experimental plan, i.e., the distribution of measurements in the input space, that will ensure the most precise identification. Experiments conducted in this fashion are called designed experiments, and the branch of statistics that studies designed experiments is referred to as design of experiments (DoE). A lot of different methodologies have appeared based on modeling and statistics to improve the general quality of manufactured goods [3].

DoE methods can be divided into two basic groups, based on their goals. The first group of method deals with a precise parameter estimation. The model structure for these kinds of DoE methods is needed beforehand in order to determine the experimental design. The second group aims at homogeneously covering the experimental space. These kinds of DoE methods do not need a predefined model and are normally used when the model is not known. Our approach is to combine both approaches, where the model evolves during the experiments and is also used to determine the experimental plan.

DoE methods can be further divided into the following:
1) *Classical experimental designs:* The samples are distributed according to a predefined pattern. These patterns are usually designed for a rectangular or spherical experimental region and are suitable for certain polynomial models. Common designs are *Full Factorial*, *Fractional Factorial*, *Central Composite*, etc. [4].
2) *Optimal experimental designs:* Given the model, the sample distribution is chosen so that it minimizes a certain optimality criterion. This criterion can be based on the variance of the model parameters or on the variance of the model output. A nonregular experimental region can be used in this case. The models are usually limited to the structures that have linear parameters. Common criteria are $D, A, E, G, I,$ and *V-optimality* [5], which will be given in more detail later.
3) *Space-filling designs:* The goal of space-filling designs is to fill the experimental region as homogeneously as possible using a limited number of samples. They do not depend on any predefined model. Common methods are *Latin Hypercube Sampling*, *Maximin,* and *Minimax* designs [6], [7].

The main disadvantage of the model-based DoE methods is the requirement for a predefined model. It is often the case that the model of the system is not known in advance and therefore needs to be identified. On the other hand, space-filling designs can be used to explore the experimental region and then determine the suitable model. The main disadvantage of space-filling designs is that they do not take into account nonlinearities when determining experimental plans. The proposed method combines the advantages of both basic approaches, starting with the methods of space-filling design and then involving the evolving structure of the fuzzy model into the experimental design plan.

In the literature, the optimal experimental design for neural networks in combination with active learning is presented in [8]. Hametner *et al.* [9] discuss the optimal experimental design for neural networks and local model networks, where dynamical systems are modeled. In [10], the minimization of the bias error rather than the parameter or output variance is discussed. When the error variance is small in comparison with the bias error, minimizing the bias error is beneficial. A method based on active

learning and hierarchical local model trees, which focuses on minimizing the model bias error, is proposed in [11] and [5].

The purpose of this contribution is to present an experimental design method based on a supervised, hierarchical clustering algorithm (SUHICLUST) for fuzzy model identification [12]. The main advantage of the SUHICLUST algorithm is that it results in a parsimonious fuzzy model. In combination with an active-learning approach, a suitable model for a linear or nonlinear system can be found with a small number of measurements.

This paper is structured as follows. In Section II, the fuzzy model in the Takagi–Sugeno (T–S) form is described, together with the supervised hierarchical clustering identification algorithm. In Section IV, the proposed method for the supervised hierarchical-clustering-based DoE is presented. Section V shows the demonstration of the proposed method. This paper is concluded in Section VIII.

II. NONLINEAR MODEL DESCRIBED IN THE TAKAGI–SUGENO FUZZY FORM

The form of the T–S [13] fuzzy model used here is given in a simplified form as follows:

$$\mathbf{R}_j : \text{ if } \boldsymbol{u}_p \text{ is } A_j, \text{ then } y = y_j(\boldsymbol{u}), \quad j = 1, \dots, m. \quad (1)$$

The q-element vector $\boldsymbol{u}_p^T = [u_{p1}, \dots, u_{pq}]$ denotes the input or variables in premise, and the variable y is the output of the model. Each input vector $\boldsymbol{u}_p$ is connected with the q-dimensional function F_j, $j = 1, \dots, m$, which defines the corresponding fuzzy set A_j, $j = 1, \dots, m$, and each fuzzy set A_j, $j = 1, \dots, m$, is associated with a real-valued function $\phi_{A_j}(\boldsymbol{u}_p) : \mathbb{R} \to [0,1]$, i.e., the membership function that produces the membership grade of the variable $\boldsymbol{u}_p$ with respect to the fuzzy set A_j.

According to the nature of the system, the variables in $\boldsymbol{u}_p$ that define the partitioning of the input–output space are not necessarily the inputs of the fuzzy system or the variables in the regression vector. Therefore, the r-element vector $\boldsymbol{u}^T = [1, u_1, \dots, u_{r-1}]$ is formed to represent the input to the fuzzy system. It is usually referred to as the consequence vector. Augmenting the unit value 1 to the input vector results in an affine fuzzy model. Depending on the use of the model the functions $y_j(\cdot)$ can be arbitrarily chosen smooth functions, but affine functions are normally used.

The system in (1) can be described in closed form if the intersection of the fuzzy sets is previously defined. The generalized form of the intersection is the so-called *triangular norm* (T-norm). In our case, the latter was chosen as an algebraic product providing the output of the fuzzy system

$$y = \frac{\phi_{A_j}(\boldsymbol{u}_p) \, y_j(\boldsymbol{u})}{\sum_{j=1}^{m} \phi_{A_j}(\boldsymbol{u}_p)}, \qquad j = 1, \dots, m. \quad (2)$$

Equation (2) can be simplified by a consideration of unity partitioning where the functions $\mu_j(\boldsymbol{u}_p)$ are defined by

$$\mu_j(\boldsymbol{u}_p) = \frac{\phi_{A_j}(\boldsymbol{u}_p)}{\sum_{j=1}^{m} \phi_{A_j}(\boldsymbol{u}_p)}, \qquad j = 1, \dots, m \quad (3)$$

which gives information about the fulfillment of the respective fuzzy rule in the normalized form.

Combining (2) and (3), the following equation is derived:

$$y = \sum_{j=1}^{m} \mu_j(\boldsymbol{u}_p) y_j(\boldsymbol{u}). \quad (4)$$

Equation (4) shows that the output of a fuzzy system is a function of the premise vector $\boldsymbol{u}_p$ (q-dimensional) and the consequence vector $\boldsymbol{u}$ (r-dimensional). The first one defines the input–output partitioning and the second one the regression vector. The dimension of the input space may be lower than $(q + r)$ since it is very common to have the same variables present in the vectors $\boldsymbol{u}_p$ and $\boldsymbol{u}$.

Usually, the output value is defined as a linear combination of the consequence states as given in

$$y_j(\boldsymbol{u}) = \boldsymbol{\psi}_j^T \boldsymbol{\theta}_j, \quad j = 1, \dots, m \quad (5)$$

where the model parameters vector with the dimension $1 \times r$ is denoted as $\boldsymbol{\theta}_j^T = [\theta_{j0}, \theta_{j1}, \dots, \theta_{j,r-1}]^T$ and where $\boldsymbol{\psi}_j^T$ with the dimension $1 \times r$ stands for the weighted augmented regression vector $\boldsymbol{\psi}_j^T = \mu_j(\boldsymbol{u}_p)\boldsymbol{u}$, $j = 1, \dots, m$.

Equation (4) consists of m local linear models and can be written as

$$y = \sum_{j=1}^{m} \boldsymbol{\psi}_j^T \boldsymbol{\theta}_j. \quad (6)$$

The whole dataset consists of n measurements, each weighted regression data vector can therefore be written as $\boldsymbol{\psi}_{jk} = \mu_{jk}\boldsymbol{u}_k$, $k = 1, \dots, n$, $j = 1, \dots, m$, and the whole data matrix $\boldsymbol{\Psi}_j$ with the dimensions $r \times n$ is written as

$$\boldsymbol{\Psi}_j = [\boldsymbol{\psi}_{j1}, \dots, \boldsymbol{\psi}_{jn}]. \quad (7)$$

The output data vector of the dimension $n \times 1$ is then written as $\boldsymbol{y}_j = [\mu_{j1}y_1, \dots, \mu_{jn}y_n]^T$, $j = 1, \dots, m$.

The weighted relation between the model input and model output at the time instant k is then written as

$$y_{jk} = \boldsymbol{\psi}_{jk}^T \boldsymbol{\theta}_j \quad (8)$$

or for the whole dataset as

$$\boldsymbol{y}_j = \boldsymbol{\Psi}_j^T \boldsymbol{\theta}_j. \quad (9)$$

The parameters of the affine fuzzy model are, in the case of batch data, obtained using the weighted least-square method (wLS) for each local model. The fuzzy model in the form given in (6) is referred to as the affine T–S model and can be used to approximate any arbitrary function with any desired degree of accuracy [14], [15].

A. Supervised Hierarchical Clustering for Fuzzy Model Identification

The main problem in the case of fuzzy identification remains the partitioning of the data space. The supervised hierarchical clustering algorithm combines the advantages of the hierarchical approach, where the input space is divided in a hierarchical manner to determine the validity functions, and the fuzzy clustering approaches, which find characteristic regions of data samples,

i.e., clusters, which correspond to the local models. The main advantage is the ability to adapt to system nonlinearities with a small number of local models, the accuracy of global and local models and the fact that the algorithm practically converges to the same results, which is not the case when using classical fuzzy clustering approaches.

In this paper, Gustafson–Kessel (GK) fuzzy clustering [16] will be combined with a hierarchical decision-tree algorithm. GK fuzzy clustering finds clusters of different shapes in the dataset $\mathbf{Z}$, which consists of the input premise data $\boldsymbol{u}_p$ and the output y. The data matrix $\mathbf{Z}$ is defined as $\mathbf{Z}^T = \left[\boldsymbol{z}_1^T, \ldots, \boldsymbol{z}_n^T \right]$, where $\boldsymbol{z}_k$ denotes the vector of measurements in the time instant k, as follows: $\boldsymbol{z}_k = [\boldsymbol{u}_{pk} \; y_k]$. It is assumed that the measured data are normalized, i.e., centered and scaled, to simplify the notations. The clusters are described by cluster centers $\boldsymbol{c}_i$, $i = 1, \ldots, m$, and the corresponding fuzzy covariance matrices $\mathbf{F_i}$, $i = 1, \ldots, m$.

The main steps of the algorithm follow the simplified version of the algorithm from [17]. They are as follows: determination of the initial T–S fuzzy model with the number of clusters $m = 2$, and determination of the worst-modeled local model defined by the index w, which corresponds to the model with the largest variation σ_j defined as

$$\sigma_{e_j}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}(\boldsymbol{u}_{pi}))^2 \mu_j(\boldsymbol{u}_{pi})}{\sum_{i=1}^{n} \mu_j(\boldsymbol{u}_{pi})}, \qquad j = 1, \ldots, m. \tag{10}$$

Next, the domain of the worst model is split into the two new clusters by applying the GK fuzzy clustering algorithm. The cluster centers are optimized again using GK, where the old and the new cluster centers are used as cluster-center prototypes. The number of clusters is therefore increased by one $m = m + 1$, and then, the determination of the validity functions and the estimation of the parameters of the local models are made. The steps from the second to the final step are repeated until the stop condition is met. The stop condition can be either the model quality or the maximum number of local models $m_{\max}$.

The pseudocode of the algorithm is given in Algorithm 1.

III. OVERVIEW OF EXPERIMENTAL DESIGN METHODS

In this section, a brief overview of experimental design methods will be given. They can be divided according to the use of the model into classical and space-division methods and optimal design methods that are model based.

A. Classical Approaches

With classical experimental design methods, the measurements are arranged based on a predefined geometrical pattern. It is assumed that the shape of an input space is either a hypercube or a hyperellipsoid. These methods were developed for polynomial approximation models. Some of the most commonly used approaches are full factorial design, fractional (or incomplete) factorial design, central composite design, etc.

The full fractional design is specified by the number of input variables n_u and the resolution r. The approach defines r

---

Algorithm 1 Pseudocode of the iterative fuzzy model identification algorithm

1: Calculation of the data center $\mathbf{c}$ and the covariance matrix $\mathbf{C} = \frac{1}{n-1}\mathbf{Z}^T\mathbf{Z}$
2: Singular value decomposition of matrix $\mathbf{C} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$
3: Define initial prototypes: $\mathbf{v}_1 = \mathbf{c} + \sqrt{\lambda_1}\mathbf{p}_1$, $\mathbf{v}_2 = \mathbf{c} - \sqrt{\lambda_1}\mathbf{p}_1$, $\mathbf{p}_1$ is the main eigenvector with eigenvalue λ_1
4: GK clustering using all the data results in cluster-centers
   $\mathbf{c}_1$ and $\mathbf{c}_2$.
5: **Repeat**
6:   Compute $\mu_j(\mathbf{u}_{pk})$, $j = 1, \ldots, m$, $k = 1, \ldots, n$.
7:   Compute $\boldsymbol{\theta}_j$, $j = 1, \ldots, m$ using wLS algorithm.
8:   Compute σ_{e_j}, $j = 1, \ldots, m$.
9:   Define the cluster with the largest σ_{e_j}, $(j = w)$.
10: Calculation of the center $\mathbf{c_w}$ and the covariance matrix
    $\mathbf{C}_w = \frac{1}{n-1}\mathbf{Z}_w^T\mathbf{Z}_w$ for the worst cluster.
11: Compute the singular value decomposition of $\mathbf{C}_w = \mathbf{P}_w\boldsymbol{\Lambda}_w\mathbf{P}_w^T$.
12: For the cluster $j = w$ define initial prototypes: $\mathbf{v}_{w1} = \mathbf{c}_w + \sqrt{\lambda_{w1}}\mathbf{p}_{w1}$, $\mathbf{v}_{w2} = \mathbf{c}_w - \sqrt{\lambda_{w1}}\mathbf{p}_{w1}$.
13: Define initial prototypes: $(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_{m+1}) = (\mathbf{c}_1, \ldots, \mathbf{c}_j, \ldots, \mathbf{c}_{m+1})$.
14: GK clustering using all the data and initialization with cluster prototype $\mathbf{v}_j$, $j = 1, \ldots, m+1$.
15: **Until** (end criterion is not met)
16: Compute $\mu_j(\mathbf{u}_{pi})$, $i = 1, \ldots, n$, $j = 1, \ldots, m$.
17: Compute $\boldsymbol{\theta}_j$, $j = 1, \ldots, m$ using wLS.
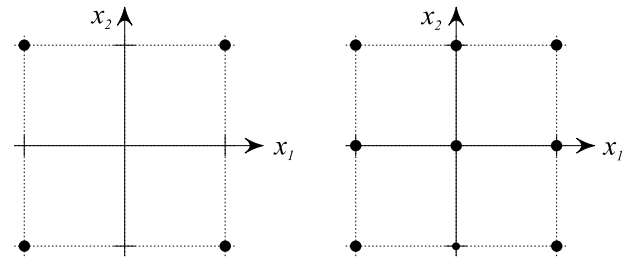18: Compute output of the model $\hat{y}(\mathbf{u}_{pk})$, $k = 1, \ldots, n$

Fig. 1. Full factorial design. Dotted lines denote the experimental region and dots the measuring points. Left: 2^{n_u}. Right 3^{n_u}.

equally spaced measuring points for each input variable. The DoE comprises all the possible point combinations. With this approach, r^{n_u} measurements are made. The usual notation for the full fractional design is r^{n_u}. The most common are 2^{n_u} and 3^{n_u} (see [6]), shown in Fig. 1. The main disadvantage of the full factorial design is the number of generated measuring points. The number increases exponentially with the dimension of the input space. With a high-dimensional input space and a high resolution, the number of measuring points often becomes too large for practical applications.

The fractional factorial design approach requires fewer measurement points than the complete factorial design. However, it also results in a less accurate model-parameter estimation.

It is usually denoted by $r^{n_u - p}$. The left-out measurements are determined based on an implemented model and method.

There are many other classical DoE approaches, such as Plackett–Burman, central composite design, and Box–Behnken. A detailed overview of classical experimental design approaches can be found in [4].

B. Space-Filling Design

Space-filling designs are experimental designs that do not use a process model. Their goal is to uniformly cover the experimental region with measuring points. They are normally used for the DoE of deterministic processes (experimenting with complex mathematical models), where there are no disturbances or noise. The only source of error is the model bias. Therefore, the goal is to determine an unbiased model. This is achieved by selecting the measuring points throughout whole experimental region (e.g., regular grid). Incorrect assumptions made in the model of the relation between the inputs and the response can produce a bias model. The space-filling designs can also be used to obtain the initial information about an unknown system. They are used to gather enough information to construct a correct model structure for further experimentation based on a process model. To obtain the first information about the system, a full factorial DoE can be used. However, the number of generated measuring points with the full factorial DoE increases exponentially with the dimensions of the experimental region. The space-filling designs tend to partition the experimental region as uniformly as possible, with the number of measuring points usually being smaller than n^k.

The most frequently used techniques for the space-filling design are *minimax*, *maximin*, and *Latin hyper cube* design. The *minimax* design minimizes the maximum distance between two measurements for the whole set of measurements in the domain as given with the criterion

$$\min_{\mathbf{U}} \max_{\boldsymbol{u}_i, \boldsymbol{u}_j \in \mathbf{U}} d(\boldsymbol{u}_i, \boldsymbol{u}_j) \tag{11}$$

where $\mathbf{U}$ defines the subset with n elements (experimental plan) of all the possible measuring points $\mathbf{D}$ (experimental region), and $d(\boldsymbol{u}_i, \boldsymbol{u}_j)$ is the distance between two measuring points $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$.

The *maximin* distance maximizes the minimum distance between two measurements for the whole set of measurements in the domain as given with the criterion

$$\max_{\mathbf{U}} \min_{\boldsymbol{u}_i, \boldsymbol{u}_j \in \mathbf{U}} d(\boldsymbol{u}_i, \boldsymbol{u}_j). \tag{12}$$

More about the *minimax* and *maximin* plans can be found in [18].

The *Latin hypercube design* (Latin hypercube sampling) divides each input variable range into n equally probable intervals. The n measurement points are then placed to satisfy the Latin hypercube requirements, meaning that when projecting measuring points on one dimension, each point falls into one interval. This way, the entire experimental region is covered. An example for $n = 5$ is shown in Fig. 2.
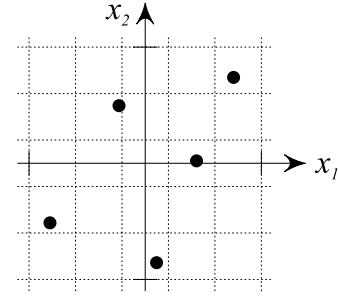

Fig. 2. Latin hypercube. $n = 5$. Dotted lines represent the intervals which divide the input space.
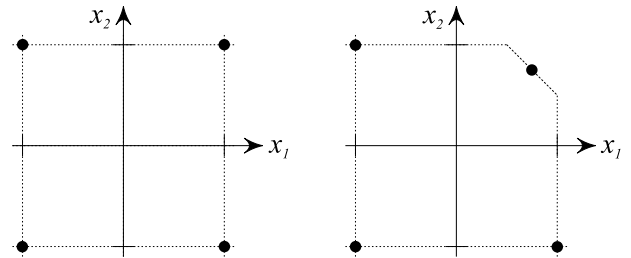

Fig. 3. Full factorial design. Dotted lines denote the experimental region and dots the measuring points. (Left) Quadratical experimental area. (Right) Constrained experimental area.

C. Optimal Experimental Design Approaches

In practice, there are a lot of problems that cannot be solved using conventional approaches due to the following limitations: the limited number of measurements, which does not correspond to the required number of measurement points calculated with the classical DoE; and the experimental region has a certain limitation. The classical DoE approaches usually require the hypercube or hyperellipsoidal experimental space. However, these requirements cannot always be met, due to real process limitations. In real experiments some input combinations are not allowed, because they would lead to process instability or a dangerous operating regime. An example is shown on the right-hand side of Fig. 3.

To overcome the described problems, each problem requires its own DoE method, which has to be adjusted to the problem. The designer has to design the DoE in the sense of a proper model structure, the definition of the experimental region, the number of allowed measurements and, most importantly, the choice of the optimization criterion.

The optimization criteria can be divided into two groups. The first group represents the criteria in the parameter space that minimize the variance of the model parameters, and the second group represents the criteria in the time domain that minimize the model-prediction variance.

1) D, A, and E-Optimal Criteria: The optimal DoE is based on an optimization of the different values of the Fisher information matrix $\mathbf{I}$. This means the optimization of the model-parameter accuracy. The information matrix is calculated from the derivative of the model output with respect to the model parameters, as follows:

$$\mathbf{I} = \frac{1}{\sigma^2} \boldsymbol{\Psi}^T \boldsymbol{\Psi} \tag{13}$$

where σ^2 represents the measurement variance, and $\mathbf{\Psi}$ is the gradient matrix of the model output

$$\mathbf{\Psi}^T = \left[\left(\frac{\partial \hat{y}(\boldsymbol{u}_1, \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \right)^T, \ldots, \left(\frac{\partial \hat{y}(\boldsymbol{u}_n, \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \right)^T \right]. \quad (14)$$

The Fischer information matrix measures the information of the input regressor vector. The matrix is real and symmetric and its inverse gives the dispersion matrix or covariance matrix of the model parameters [9].

The optimization of the information matrix means better excitation of the input variables and consequently better conditioning in the case of identification. To optimize the Fischer matrix, one of the following criteria can be chosen. The *D-optimal criterion*, which minimizes the determinant of the covariance matrix and maximizes the information matrix. The criterion results in minimizing the generalized variance of the parameter estimates for a prespecified model. Lowering the parameter variance increases the accuracy of the estimated parameters. The use of *the D-optimal criterion* is very common and can be defined as $J_D = \max |\mathbf{\Psi}^T \mathbf{\Psi}|$. *The A-optimal criterion* minimizes the trace of the parameter covariance matrix. The criterion results in minimizing the average variance of the model parameters and is given as $J_A = \min \left(trace \left(\mathbf{\Psi}^T \mathbf{\Psi} \right) \right)$. *The E-optimal criterion* maximizes the smallest eigenvalue of the information matrix, as given by $J_E = \max \left(\lambda_{\min} \left(\mathbf{\Psi}^T \mathbf{\Psi} \right) \right)$.

2) G, I, and V-Optimal Criteria: Sometimes the problem requires us to optimize the accuracy of the model prediction. The criteria presented next minimize the variance of the model prediction, and with that maximize the model prediction accuracy. They can use different criterion functions, such as maximum prediction variance or average prediction variance. The most common criterion is *the G-optimal criterion*, which minimizes the maximum model-prediction variance over the experimental region, which is given as

$$J_G = \max_{\boldsymbol{u}} (\text{var}(\hat{y}(\boldsymbol{u}))), \ \boldsymbol{u} \in \boldsymbol{T}$$

where $\mathbf{T}$ denotes the experimental region. *The I-optimal criterion* minimizes the average model-prediction variance over the experimental region, as follows: $J_I = \min \left(\frac{1}{A} \sum_{\mathbf{T}} \text{var}(\hat{y}(\boldsymbol{u})) \right)$, where A stands for the experimental region volume, and *the V-optimal criterion* minimizes the average model-prediction variance at certain measuring points, as follows:

$$J_V = \min \left(\frac{1}{n} \sum_{i=1}^{n} \text{var}(\hat{y}(\boldsymbol{u}_i)) \right), \quad \boldsymbol{u}_i \in \boldsymbol{T}.$$

The most commonly used are *the D and I-optimal criteria* [4]. The drawback of the optimal DoEs is, the same as with the classical DoEs, that they neglect the bias. They assume that the models are unbiased [5], [10]. In the case of disturbances and unbiased models, the optimal DoEs are very efficient. Some examples can be found in [19].

3) Optimal Design of Experiments for Nonlinear Systems: The optimal DoE methods were first developed for polynomial models; however, they can be used for all types of models that have linear parameters. There are several successful implementations of the optimal DoE in combination with neural networks and fuzzy models. In [8], an optimal criterion for a pattern-recognition neural network is proposed. This criterion is extended in [20] for dynamic neural networks. The approach is based on an analytic optimization of the Fischer information matrix; therefore, the input selection does not depend on the candidate set. In [9], an approach to the analytic optimization of the information matrix for dynamic systems is shown. The DoE approach considers the process output limitations and the dynamic limitations of the inputs. The approach was tested for dynamic neural networks and fuzzy models.

IV. DESIGN OF EXPERIMENT BASED ON SUPERVISED HIERARCHICAL CLUSTERING

The SUHICLUST-based DoE algorithm iteratively adds samples to the dataset $\mathbf{Z}$ and models the dataset with T–S fuzzy model of increasing complexity. At each iteration, samples are added in the region of the worst fuzzy model, which implicitly indicates the system's nonlinearity.

The assumptions when using DoE based on supervised hierarchical clustering are as follows: The fuzzy model structure is unknown; the knowledge of the process is limited (only the experimental region is known in advance); disturbances and noise are limited; and the main source of model error is the model's inadequate structure.

The proposed DoE is based on the supervised hierarchical clustering presented in Section II-A. The clustering method was originally an iterative method. Therefore, it is easy to transform it into an online clustering method. This is done in the following way: First, the fuzzy model is learned on a small amount of data; then, the measurements are made in the experimental region where the model's performance does not satisfy a certain criterion. At each iteration, the model structure is evolved and its parameters are adapted considering the new measurements. The SUHICLUST framework was chosen as it is able to approximate a nonlinear process well with a small number of local models. Therefore, fewer measuring points are needed to identify the model.

A. Supervised Hierarchical Clustering Algorithm Design of Experiment Procedure

A more detailed explanation of the SUHICLUST-DoE procedure will be given next. The procedure consists of initialization, calculation of the initial fuzzy model parameters, design of the experimental plan, update of the fuzzy model parameter, and evaluation of the current fuzzy model.

1) Initialization: The learning procedure starts with the initial set of data $\mathbf{Z}_o = \left[\boldsymbol{z}_1^T, \ldots, \boldsymbol{z}_{n_o}^T \right]^T$ with the dimensions $n_o \times (q + 1)$, where $\boldsymbol{z}_k = [\boldsymbol{u}_{pk} \ y_k]$, $k = 1, \ldots, n_o$, and n_o stands for the number of measurements in the initial dataset. The initial set of data should be evenly distributed across the whole experimental domain obtained using the Latin hypercube approach, full factorial approach, etc.

The number of data n_o in the initial set must be large enough to estimate the parameters of two affine local models. The amount of initial data is defined *ad hoc* and depends on the nature of the nonlinearity and the complexity of the problem.

The parameters of the method that should be defined are the number of new measuring points n_w and the parameter α that defines the implemented fuzziness (usually set $\alpha = 2$). Both parameters $l_{\max}$ and n_w implicitly define the final number of measurements, which is then $n_o + l_{\max} n_w$. Next, the termination criterion defined as the number of maximal iterations $l_{\max}$ or the maximal allowed normalized root-mean-squared error ($nrmse$) in (15) should also be defined

$$nrmse = \sqrt{\frac{\sum_{k=1}^{n}(y(\mathbf{u}_k) - \hat{y}(\mathbf{u}_k))^2}{\sum_{k=1}^{n}(y(\mathbf{u}_k) - \bar{y})^2}}, \quad \bar{y} = \frac{1}{n}\sum_{k=1}^{n}y(\mathbf{u}_k). \tag{15}$$

2) Initial Fuzzy Model Parameters: Based on the initial dataset $\mathbf{Z}_o$ the initial fuzzy model with two local models ($m = 2$) is defined. First, two cluster centers are positioned in the main direction of the data spread [12]. The direction is obtained using an eigenvalue decomposition of the data covariance matrix $\mathbf{C}_o$,

$$\mathbf{C}_o = \frac{1}{n_o - 1}\mathbf{Z}_o^T \mathbf{Z}_o = \mathbf{P}_o \mathbf{\Lambda}_o \mathbf{P}_o^T \tag{16}$$

where $\mathbf{P}_o$ represents the matrix of eigenvectors, and $\mathbf{\Lambda}_o$ represents a diagonal matrix of eigenvalues. The initial position of the first two cluster centers is defined as

$$\boldsymbol{v}_{o1} = \boldsymbol{c}_o + \sigma_{o1}\boldsymbol{p}_{o1}, \qquad \boldsymbol{v}_{o2} = \boldsymbol{c}_o - \sigma_{o1}\boldsymbol{p}_{o1} \tag{17}$$

where σ_{o1} is the square root of the maximum eigenvalue λ_{o1} ($\sigma_{o1} = \sqrt{\lambda_{o1}}$), and $\boldsymbol{p}_{o1}$ represents the corresponding eigenvector.

The optimal positions of the cluster centers $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ and their fuzzy covariance matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ are then calculated using the GK method

$$(\{\boldsymbol{c}_1, \boldsymbol{c}_2\}, \{\mathbf{F}_1, \mathbf{F}_2\}) \leftarrow \mathrm{GK}(\mathbf{Z}_o, \{\boldsymbol{v}_1, \boldsymbol{v}_2\}, \alpha). \tag{18}$$

To obtain the normalized membership values $\mu_j(\boldsymbol{u}_{pk})$, first, the distance d_{jk}^2 from the data point defined by the index k to each cluster center ($j = 1, 2$) is calculated, taking into account the fuzzy covariance matrix $\mathbf{F}_j$, $j = 1, 2$, which scales and rotate the axes

$$d_{jk}^2 = (\boldsymbol{z}_k - \bar{\boldsymbol{c}}_j)^T \bar{\mathbf{F}}_j^{-1}(\boldsymbol{z}_k - \bar{\boldsymbol{c}}_j), \quad j = 1, 2, \; k = 1, \ldots, n_o \tag{19}$$

where $\bar{\boldsymbol{c}}_j$, $j = 1, 2$, and $\bar{\mathbf{F}}_j$, $j = 1, 2$, stand for the projections of the cluster center $\boldsymbol{c}_j$, $j = 1, 2$, and the fuzzy covariance matrix $\mathbf{F}_j$, $j = 1, 2$, to the input data space domain as follows:

$$\bar{\boldsymbol{c}}_j = \boldsymbol{c}_j \mathbf{I}_{\mathrm{proj}}, \quad j = 1, 2 \tag{20}$$

$$\bar{\mathbf{F}}_j = \mathbf{F}_j \mathbf{I}_{\mathrm{proj}}, \quad j = 1, 2 \tag{21}$$

where $\mathbf{I}_{\mathrm{proj}}$ stands for

$$\mathbf{I}_{\mathrm{proj}} = \begin{bmatrix} \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{22}$$

and $\mathbf{I}_q$ denotes the unit matrix of dimension $q \times q$.

The membership values $\phi_j(\boldsymbol{u}_{pk})$, $j = 1, \ldots, 2$, $k = 1, \ldots, n_o$, are calculated by

$$\phi_j(\boldsymbol{u}_{pk}) = e^{-\gamma d_{jk}^2} \tag{23}$$

where $\gamma \geq 1$ is a factor that defines the smoothness of the Gaussian function. The normalized membership value is then defined as

$$\mu_j(\boldsymbol{u}_{pk}) = \frac{\phi_j(\boldsymbol{u}_{pk})}{\sum_{j=1}^{2}\phi_j(\boldsymbol{u}_{pk})}, \quad j = 1, 2, \; k = 1, \ldots, n_o. \tag{24}$$

At the end of the fuzzy model initialization, the local model parameters $\boldsymbol{\theta}_j$, $j = 1, \ldots, 2$, are estimated using the wLS method.

3) Design of the Experimental Plan: In this step, the fuzzy model for the existing dataset is evaluated. Based on its performance, the experimental plan is determined. Using the performance analysis, we determine the worst local model. New measurement points are placed in the cluster area with the worst local model to provide new information to improve the local model's performance.

The evaluation of the local fuzzy model is estimated using the normalized sum of the squared error

$$\sigma_j^2 = \frac{\sum_{k=1}^{n}(y_k - \hat{y}(\boldsymbol{u}_{pk}))^2 \mu_j(\boldsymbol{u}_{pk})}{\sum_{k=1}^{n}\mu_j(\boldsymbol{u}_{pk})}, \quad j = 1, \ldots, m. \tag{25}$$

The worst local model is denoted with the index w, which is defined as follows:

$$w = \arg\max_j \sigma_j^2. \tag{26}$$

The new experimental region is defined by the worst-performing local model w. The possible inputs are all input vectors $\boldsymbol{u}_p$ that satisfy the following criterion:

$$\mu_w(\boldsymbol{u}_p) > \delta \tag{27}$$

where the parameter δ defines the width of the new experimental region. The usual value of $\delta = 0.5$.

The next step is the design of the experimental plan. In general, there are several possible approaches where the methods of space-filling design are used. In our case, we chose the design approach that includes the information from previous measuring points in the current area. This way, we get more information about the whole region of the worst local model. First, we generate a set of random measuring points that satisfy the condition in (27). Next, we choose n_w measuring points in such a way that the new and old measuring points are uniformly distributed. This is done using the pseudo-Monte Carlo sampling algorithm (PM-CSA) [5]. The positions of the selected measuring points are chosen using the maximin optimization criterion given in (12).

The PM-CSA algorithm calculates the distances between all existing measuring points and the newly generated candidates for the measuring points. For each existing point, the set of neighborhood points from the candidate set is selected. From the neighborhood points, a point farthest from the existing measuring point is selected as a new measuring point. The procedure is repeated until all the n_w new measuring points are selected. The main advantage of this algorithm is that the selected points
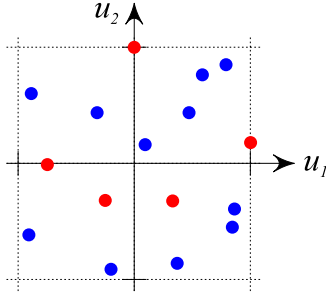
Fig. 4. Example of the experimental plan. Blue dots represent the existing measuring points, and red dots represent the new measuring points.

Fig. 5. System. (Left) Centered and normalized dataset. (Right) Final fuzzy model after eight iterations.

uniformly cover the region, taking into account the existing previous measuring points.

An example of the generated experimental plan with the described method is shown in Fig. 4. It is clear that the new measuring points fill the region where no old measuring points are present.

4) Updating the Fuzzy Model: When a new measurement set is obtained for the experimental region of the worst local model, the cluster is split into two clusters. This is done in a similar way as with the algorithm initialization. When defining the main spread direction, only the data belonging to the wth local model (cluster) are used. The new dataset with the old measurements and the newly generated is denoted as $\mathbf{Z}_w$. The number of measurements in $\mathbf{Z}_w$ is now equal to n_{nw}, where $n_{nw} = n_{ow} + n_w$. The eigenvectors and eigenvalues are calculated as

$$\mathbf{C}_w = \frac{1}{n_w - 1}\mathbf{Z}_w^T\mathbf{Z}_w = \mathbf{P}_w\mathbf{\Lambda}_w\mathbf{P}_w^T. \qquad (28)$$

The cluster is split into two parts, and the new center locations are defined as

$$\begin{aligned}\boldsymbol{v}_{w1} &= \boldsymbol{c}_w + \sigma_{w1}\boldsymbol{p}_{w1} \\ \boldsymbol{v}_{w2} &= \boldsymbol{c}_w - \sigma_{w1}\boldsymbol{p}_{w1}\end{aligned} \qquad (29)$$

where σ_{w1} stands for the square root of the maximum eigenvalue λ_{w1} ($\sigma_{w1} = \sqrt{\lambda_{w1}}$), and $\boldsymbol{p}_{w1}$ represents the corresponding eigenvector.

The initial center positions $\{\boldsymbol{v}_{w1}$ and $\boldsymbol{v}_{w2}\}$ in (29) are used to apply the local GK clustering algorithm on the dataset $\mathbf{Z}_w$ as follows:

$$(\{\boldsymbol{c}_{w1}, \boldsymbol{c}_{w2}\}, \{\mathbf{F}_{w1}, \mathbf{F}_{w2}\}) \leftarrow \mathrm{GK}\left(\mathbf{Z}_w, \{\boldsymbol{v}_{w1}, \boldsymbol{v}_{w2}\}, \alpha\right). \qquad (30)$$

Two new centers $\boldsymbol{c}_{w1}$ and $\boldsymbol{c}_{w2}$ and the old centers are then used in the global GK clustering on the whole dataset $\mathbf{Z}$

$$\{\boldsymbol{c}_j, \ j = 1, \ldots, m+1\}, \{\mathbf{F}_j, \ j = 1, \ldots, m+1\}$$
$$\leftarrow \mathrm{GK}\left(\mathbf{Z}, \{\boldsymbol{v}_j, \ j = 1, \ldots, m+1\}, \alpha\right) \qquad (31)$$
$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{w1}, \boldsymbol{v}_{w2}, \ldots, \boldsymbol{v}_{m+1}\}$$
$$\leftarrow \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_j, , \ldots, \boldsymbol{c}_{m+1}\}. \qquad (32)$$

The number of local models is now increased to $m + 1$.

In certain regions, it happens that the clustering is made on a relatively small set of data measurements. This can produce
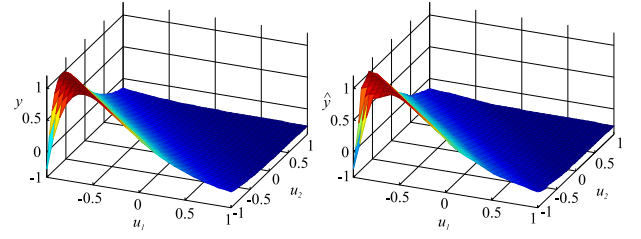
singular clusters, meaning that the surface of the cluster projection onto the input space equals zero. This must be tested before the new local model is added to the structure. If the projected fuzzy covariance matrix $\mathbf{F}_j$ is ill-conditioned, the cluster is singular and is therefore removed from the model structure. The matrix is ill-conditioned if the ratio between its highest and lowest singular value is larger than the factor β, which is usually set as 10^5. The condition is checked for each cluster. When this condition is met for a certain cluster, the cluster or local model is removed from the fuzzy model.

The last step of the iteration is to calculate the membership functions $\mu_j(\boldsymbol{u}_{pk}), j = 1, \ldots, m, \ k = 1, \ldots, n$, as proposed in (19), (23), and (24), and the local model parameters are calculated with wLS method.

The procedure continues with the next iteration (experiment design step) until the maximum number of iterations is reached ($l < l_{\max}$) or until the criterion of the optimal design is reached ($\boldsymbol{nrmse} < \boldsymbol{nrmse}_{\max}$).

The pseudocode of the DoE based on hierarchical clustering (SUHICLUST-DoE) is shown under Algorithm 2.

V. EXAMPLES OF DESIGN OF EXPERIMENTS

This section presents experiments that demonstrate the use of the proposed algorithm of fuzzy-model-based DoE [21], different design plans are compared, and the comparison of different DoE methods is given for a pharmaceutical modeling problem.

A. Demonstration of Fuzzy Design of Experiments and Comparison of Different Design Plans

The first example is given by the data that are generated using

$$y(u_1, u_2) = e^{(-(u_1+u_2))}(u_1 + u_2), \qquad u_i \in [0 \ 5], \ i = 1, 2. \qquad (33)$$

The validation dataset $\mathbf{Z}_{\mathrm{val}} = [\boldsymbol{u}_1 \ \boldsymbol{u}_2 \ \boldsymbol{y}]$ represents the set of all the potential measuring points. It is generated using a full factorial plan of dimension 50^2, which means that the number of all the measurements equals $n = 2500$. The measurement inputs u_1 and u_2 are in the interval $[0, 5]$. Before starting the algorithm, the measuring points are centered and normalized, which means the inputs are now in the interval $[-1, 1]$.

The function from (33) is in Fig. 5 on the left-hand side. To test the identification properties and the capability of the proposed algorithm, noise is added to the process output $\boldsymbol{y}$ so that the signal-to-noise ratio is 10 dB. The noisy dataset is shown in Fig. 6. The sum of the squared errors (sse), given in (34),

Algorithm 2 SUHICLUST-DoE

1: **Initialization**: Define the initial dataset $\mathbf{Z}_o$, the number of new measurements at each measurement step n_w and the maximal number of iterations $l_{\max}$ or the optimal criteria $\boldsymbol{nrmse}_{\max}$ to stop the algorithm, and the fuzziness parameter α.

2: Calculation of the data center $\mathbf{c}_o$ and the covariance matrix $\mathbf{C}_o = \frac{1}{n-1}\mathbf{Z}_o^T\mathbf{Z}_o$

3: Singular value decomposition of matrix $\mathbf{C}_o = \mathbf{P}_o\mathbf{\Lambda}_o\mathbf{P}_o^T$

4: Define initial prototypes: $\mathbf{v}_{o1} = \mathbf{c}_o + \sqrt{\lambda_{o1}}\mathbf{p}_{o1}$, $\mathbf{v}_{o2} = \mathbf{c}_o - \sqrt{\lambda_{o1}}\mathbf{p}_{o1}$, $\mathbf{p}_{o1}$ is the main eigenvector with eigenvalue λ_{o1}

5: GK clustering on dataset $\mathbf{Z}_o$ with the initial cluster centers $\mathbf{c}_{o1}$ and $\mathbf{c}_{o2}$.

6: **Repeat**

7:    Compute $\mu_j(\mathbf{u}_{pk})$, $i = 1, ..., n$, $j = 1, ..., m$.

8:    Compute $\boldsymbol{\theta}_j$, $j = 1, ..., m$ using wLS algorithm.

9:    Compute σ_{e_j}, $j = 1, ..., m$.

10:   Define the cluster with the largest σ_{e_j}, $(j = w)$.

11:   Define n_w new measurement points using pseudo-Monte Carlo sampling algorithm which satisfy $\mu_w(\mathbf{u}_p) > \delta$ and $\max_{\mathbf{U}} \min_{\mathbf{u}_i, \mathbf{u}_j \in \mathbf{U}} d(\mathbf{u}_i, \mathbf{u}_j)$.

12:   Computation of the center $\mathbf{c}_w$ and the covariance matrix $\mathbf{C}_w = \frac{1}{n-1}\mathbf{Z}_w^T\mathbf{Z}_w$ for the worst cluster with added new measurements.

13:   Compute the singular value decomposition of $\mathbf{C}_w = \mathbf{P}_w\mathbf{\Lambda}_w\mathbf{P}_w^T$.

14:   For the cluster $j = w$ compute initial prototypes: $\mathbf{v}_{w1} = \mathbf{c}_w + \sqrt{\lambda_{w1}}\mathbf{p}_{w1}$, $\mathbf{v}_{w2} = \mathbf{c}_w - \sqrt{\lambda_{w1}}\mathbf{p}_{w1}$.

15:   Compute initial prototypes: $(\mathbf{v}_1, ..., \mathbf{v}_w, \mathbf{v}_{w+1}, ..., \mathbf{v}_m) = (\mathbf{c}_1, ..., \mathbf{c}_{w1}, \mathbf{c}_{w2}, ..., \mathbf{c}_m)$.

16:   Compute GK $(\cdot)$ on the dataset $\mathbf{Z}_c$ with initial prototype $\mathbf{v}_j$, $j = 1, ..., m$, $m = m + 1$.

17:   Eliminate clusters with ill-conditioned fuzzy covariance matrices.

18:   Compute $\mu_j(\mathbf{u}_{pk})$, $k = 1, ..., n_c$, $j = 1, ..., m$.

19:   Compute $\boldsymbol{\theta}_j$, $j = 1, ..., m$ using wLS.

20:   Compute output of the model $\hat{y}(\mathbf{u}_{pk})$, $k = 1, ..., n_c$.

21: **Until** (end criterion is not met, $l < l_{\max}$ or $\boldsymbol{nrmse} < \boldsymbol{nrmse}_{\max}$)
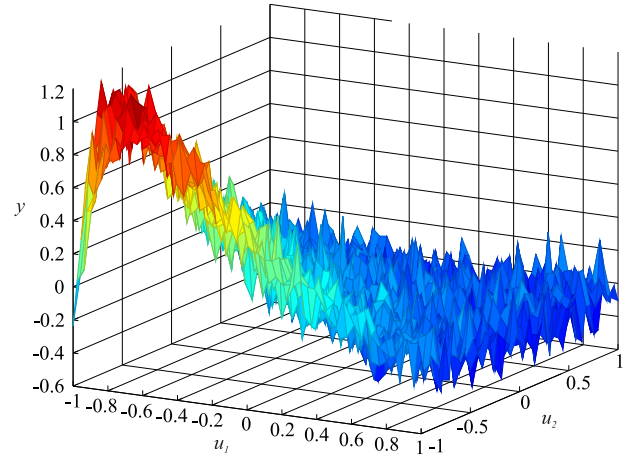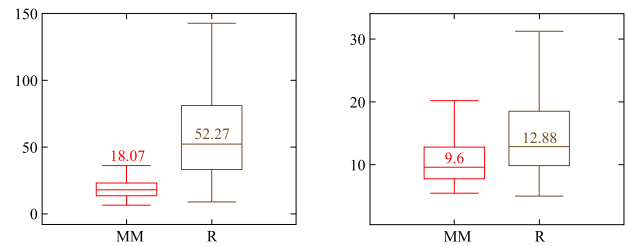

Fig. 6.   Model data with added noise.

Fig. 7.   Box plots. The numbers show the value of the median. (Left) Comparison of maximin experimental plans and random experimental plan for $n_w = 4$. (Right) Comparison of maximin experimental plans and random experimental plan for $n_w = 8$.

shown below, is used as a measurement set for comparing the performance of different experimental plans. It is calculated as the sum of the squared differences between the process output from the validation dataset $\mathbf{Z}_{\text{val}}$ and the approximated model output for all the possible input points $\boldsymbol{u}$

$$ \boldsymbol{sse} = \sum_{j=1}^{n} (y(\mathbf{u}_j) - \hat{y}(\mathbf{u}_j))^2. \qquad (34) $$

The local experimental plans in the proposed fuzzy-model-based DoE algorithm can be either *a random plan (R)*, where in each iteration in the worst local model n_w measurement points that satisfy the criterion $\mu_w(\boldsymbol{u}_p \geq 0.5)$ are randomly chosen, or *a quasimaximin plan (MM)*, where in the worst local model

n_w, measurement points are defined such that they satisfy the criterion $\mu_i(\boldsymbol{u}_p \geq 0.5)$ and the maximin criterion given in (12). They can be found using different optimization methods.

The proposed experimental plans are tested to investigate which one is the most appropriate for inclusion in the proposed DoE approach. For each experimental plan, 100 different runs are made to generate the statistics. Each plan is also tested with a different number of new measurements n_w in the local model, $n_w^1 = 4$ and $n_w^2 = 8$. The maximum number of local models is limited to 9. The maximum number of measurement points is then $n = n_w^1 \cdot m = 4 \cdot 9 = 36$ in the first example and $n = n_w^2 \cdot m = 8 \cdot 9 = 72$ in the second example.

The results are shown in Fig. 7 for $n_w^1 = 4$ on the left hand, and for $n_w^2 = 8$ measurement points on the right-hand side. The upper and lower lines of the box plot represent the maximum and minimum of the **sse**. The upper and lower edges of the box represent the first and third quartiles. The line inside the box represents the median value. The points outside the box represent the outliers.

Fig. 7 shows that, as expected, the maximin plan, with evenly distributed points in the experimental space, gives better results than those obtained with the pure randomized approach. The normal distribution of data is less sensitive to the noise. The difference is much larger, especially when the number of new measurements is smaller.

The maximin experimental plan is then used in combination with the SUHICLUST-DoE approach on the dataset defined in (33). The initial learning dataset is determined as the Latin hypercube approach with the initial number of measurements
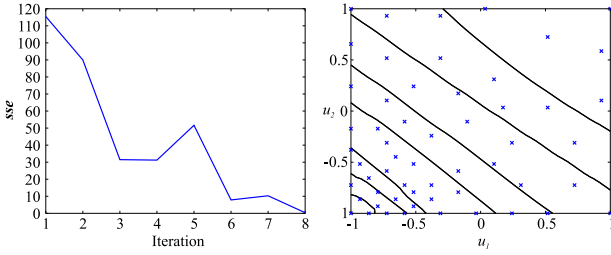
Fig. 8. Results for the simulation example. (Left) **sse** behavior. (Right) Final fuzzy model structure and measuring points.

TABLE I
COMPARISON BETWEEN DIFFERENT DoE METHODS

Method	m	n_{end}	**nrmse**
SUHICLUST	10	66	0.09
FMID	15	96	0.10
ANFIS	16	102	0.09
LOLIMOT	24	150	0.10

equal to $n_o = 14$. The maximum number of iterations is set to 8 ($l_{\max} = 8$) and the number of new measurement points is set to 6 ($n_w = 6$). As mentioned above, the quasi-maximin design plan is used to add the n_w new measurements in each iteration, which all fulfill the maximin criterion from (12).

On the left-hand side of Fig. 8, the **sse** is shown. The **sse** value is lower with every iteration. The only exception is the fifth iteration, where the **sse** increases. This phenomenon can be explained by a local unadequate fuzzy model structure. In the next iterations, the structure is improved based on the new measurements.

B. Comparison of Different Design of Experiments Methods

In order to evaluate the proposed method, a comparison of the different DoE methods applied to a pharmaceutical dataset was made. The absorption spectra of the protonation equilibria of Silychristin with a dependence on different wavelengths u_1 [nm] and pH (u_2 [pH]) at the temperature of $25\,^{\circ}$C was used in our study. This problem is presented in more detail in [22], where the noise ($\sigma_n = 0.01$) is added to the simulated data. The number of the initial learning dataset, created with the Latin hypercube plan, equals 12 and the number of new measurement points at each iteration equals 6. The comparison is given in Table I, where our method is compared with the FMID [23], ANFIS [24], and LOLIMOT [25] methods. In Table I, m stands for the number of the local models and n_{end} is the number of the measured data points. The end criterion was given with **nrmse** ≤ 0.10. The results show that the proposed method gives comparable results with a smaller number of local models. An extensive study and the main advantages of the proposed basic SUHICLUST method are given in [12].

VI. DESIGN OF EXPERIMENTS FOR THE MATHEMATICAL SIMULATION MODEL OF A LOW-NOISE AMPLIFIER

Low-noise amplifiers (LNAs) are electronic devices found at the front-end of a receiver. They are used to amplify very weak signals such as the signals received from an antenna. It is essential that the LNA has a very low signal-to-noise ratio, adds little distortion to the signal, and has a high gain so that the retrieval of this signal is possible for latter processing.

The main parameters that describe the LNA are the gain, input impedance, signal-to-noise ratio, and energy consumption. These parameters are obtained through a computer simulation of the LNA, which considers all the physical properties of the LNA's components. A simulation of this kind of system is very time consuming; therefore, the number of simulations that need to be made should be as small as possible.

To model the LNA, the DoE based on supervised hierarchical clustering is used. The main idea behind this is that the circuit designer will have an overview of the parameters' influences on the circuit's behavior. Using this kind of approach the circuit-design procedure is simplified.

For the purpose of the SUHICLUST-DoE method investigation, the mathematical model of the low-noise amplifier was taken from [26], [27]. This model will be used instead of the real LNA to simulate the measurements; however, this has no influence on our conclusions, since the use of the approach is general.

The input and output currents of the LNA were modeled using the following equations:

$$\sqrt{\overline{i_{\mathrm{in}}^2}} = \sqrt{A_{\mathrm{in}} - B_{\mathrm{in}}} \tag{35}$$

$$\sqrt{\overline{i_{\mathrm{out}}^2}} = \sqrt{A_{\mathrm{out}} - B_{\mathrm{out}}} \tag{36}$$

where

$$A_{\mathrm{in}} = |f_{gs,\mathrm{in}}|^2\, \overline{i_{gs}^2} + |f_{ds,\mathrm{in}}|^2\, \overline{i_{ds}^2}$$

$$B_{\mathrm{in}} = 2\mathrm{Im}(0.4 f_{gs,\mathrm{in}} f_{ds,\mathrm{in}}^\star)\sqrt{\overline{i_{gs}^2}\,\overline{i_{ds}^2}} \tag{37}$$

$$A_{\mathrm{out}} = |f_{gs,\mathrm{out}}|^2\, \overline{i_{gs}^2} + |f_{ds,\mathrm{in}}|^2\, \overline{i_{ds}^2}$$

$$B_{\mathrm{out}} = 2\mathrm{Im}(0.4 f_{gs,\mathrm{out}} f_{ds,\mathrm{out}}^\star)\sqrt{\overline{i_{gs}^2}\,\overline{i_{ds}^2}} \tag{38}$$

and

$$f_{gs,\mathrm{in}} = \frac{1 + j\omega L_s g_m}{1 - \omega^2 C_{gs}(L_s + L_m) + j\omega L_s g_m} \tag{39}$$

$$f_{ds,\mathrm{in}} = \frac{\omega^2 C_{gs} L_s}{1 - \omega^2 C_{gs}(L_s + L_m) + j\omega L_s g_m} \tag{40}$$

$$f_{gs,\mathrm{out}} = \frac{-j\omega(L_s + L_m) g_m}{1 - \omega^2 C_{gs}(L_s + L_m) + j\omega L_s g_m} \tag{41}$$

$$f_{ds,\mathrm{out}} = \frac{1 - \omega^2 C_{gs}(L_s + L_m)}{1 - \omega^2 C_{gs}(L_s + L_m) + j\omega L_s g_m}. \tag{42}$$

The variables in the upper equations are described in Tables II and III. In the first table, the main input variables are presented. These are the variables that can be determined by the circuit designer. Their values are determined as a function of the proper normalized values. The normalized variables W_n, L_{sn}, f_n, L_n, V_{gtn}, and L_{mn} are then from the interval $[-1, 1]$.

In the second table, the output variables that depend on the main input variables are presented.

TABLE II
MAIN INPUT VARIABLES OF THE LOW-NOISE AMPLIFIER

Var.	Meaning	Value	Unit
W	trans. width	$100 \cdot 10^{-6} \cdot 10^{W_n}$	m
L_s	source induct.	$0.5 \cdot 10^{-9} \cdot 10^{L_{sn}}$	H
f	frequency	$(11 + 10 f_n) \cdot 10^9$	Hz
L	trans. length	$(90 + 30 L_n) \cdot 10^{-9}$	m
V_{gt}	voltage	$0.275 + 0.2 V_{gtn}$	V
L_m	inductance	$1 \cdot 10^{-9} \cdot 10^{L_{mn}}$	H

TABLE III
DERIVED VARIABLES OF THE LOW-NOISE AMPLIFIER

Var.	Meaning	Value	Unit
ω	angular freq.	$2\pi f$	Hz
g_m	trans. conduct.	$1 \cdot 10^{-4} \frac{W}{L} V_{gt}$	$\frac{A}{V}$
C_{gs}	gate-source capacit.	$0.01 \cdot W L$	F
$\overline{i_{gs}^2}$	noise sp. den. (gate)	$2 \cdot 10^{-3} \frac{W}{L} V_{gt}$	$\frac{pA^2}{Hz}$
$\overline{i_{ds}^2}$	noise sp. den. (drain)	$0.5 \frac{W}{L} V_{gt}$	$\frac{pA^2}{Hz}$

TABLE IV
EXPERIMENT PARAMETERS

d_{in}	n_o	n_w	l_{max}	n_{end}	n_{test}
2	12	6	15	102	900
3	42	15	15	267	27000
4	144	36	15	684	810000

A. Experiment

The experiments were made for three different input dimensions. The parameters of the SUHICLUST-DoE algorithm used for each experiment are shown in Table IV. The number of initial measurements and the measurements per iteration are exponentially increased with the number of dimensions. The number of the initial learning dataset, created with the Latin hypercube plan, is defined as the closest integer value of expression $(\sqrt{12})^{d_{in}}$ and the number of new measurement points equals the integer number, closest to the $(\sqrt{6})^{d_{in}}$. This way of determining the number of measurements is used in order to make a fairer comparison of the algorithm for different numbers of dimensions. The number of iterations l_{max} or the number of local model m is the same in all the experiments and given as 15. The final number of measured points n_{end} for each learning example and the validation datasets are shown in Table IV, along with the number of dimensions. In the case of a 2-D experiment, the normalized input variables W_n and L_{sm} are varied. The other normalized variables are set to zero. In the 3-D experiment, the normalized input variables W_n, L_{sn}, and L_{mn} are varied, and in the 4-D experiment, the normalized input variables W_n, L_{sn}, L_{mn}, and L_n are the input variables. All the other normalized input variables are set to zero.

The learning and validation dataset is made on a grid defined by 30 evenly distributed points per dimension. The validation dataset is obtained by shifting the learning grid so that the samples did not coincide with the training samples. This way, the measurement of the mean square error (mse) $mse =$
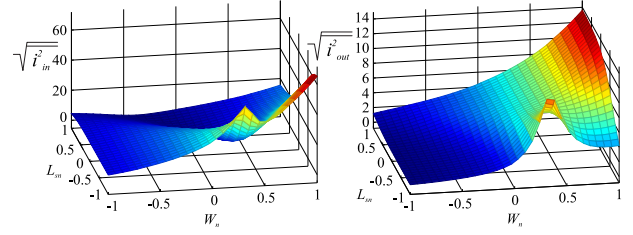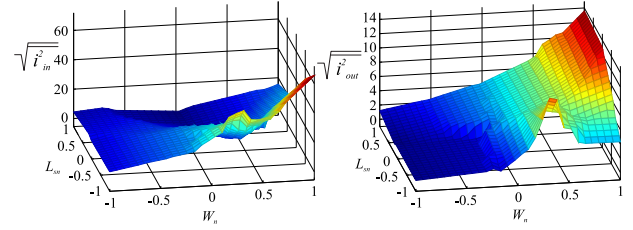

Fig. 9. Noise currents of the low-noise amplifier.

Fig. 10. Results for 2-D problem ($L_{mn} = 0$, $L_n = 0$).
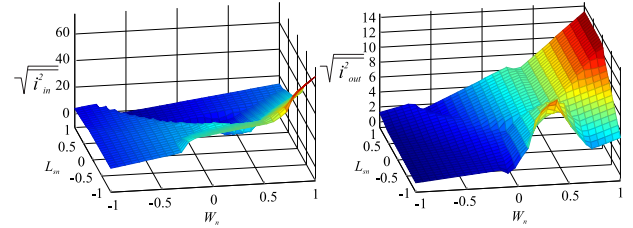

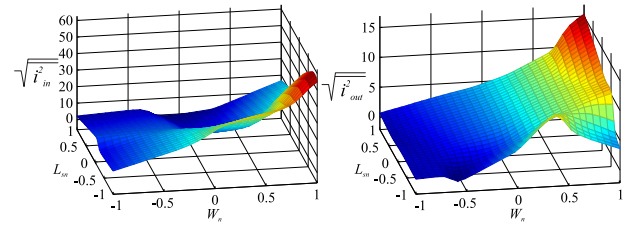
Fig. 11. Results for 3-D problem ($L_{mn} = 0$, $L_n = 0$).

Fig. 12. Results for 4-D problem ($L_{mn} = 0$, $L_n = 0$).

TABLE V
MODELING RESULTS FOR $\sqrt{i_{in}^2}$

d_{in}	mse	$nrmse$	n_{end}/n_{test}
2	2.41	0.09	0.1133
3	4.91	0.15	0.0099
4	6.44	0.21	0.0008

$\frac{1}{n} \sum_{j=1}^{n} \left(y(\mathbf{u}_j) - \hat{y}(\mathbf{u}_j) \right)^2$ is more relevant. The dataset in our problem for the 2-D experiment is shown in Fig. 9. Besides the mse, the $nrmse$ is also calculated.

VII. RESULTS AND DISCUSSION

The results are shown in Figs. 10–12 and Tables V and VI. The results for the 3-D case are plotted with L_{nm} set to zero and for the 4-D case with L_{nm} and L_n set to zero.

高

TABLE VI

MODELING RESULTS FOR $\sqrt{i_{\text{out}}^2}$

d_{in}	mse	$nrmse$	$n_{\text{end}}/n_{\text{test}}$
2	0.05	0.07	0.1156
3	0.26	0.17	0.0098
4	0.50	0.23	0.0008

From a visual comparison of the results with the learning datasets, we can see that the models describe the main characteristic of the noise current in all the cases quite well. For higher dimensions, the model is a little less accurate, but this is due to the relatively small number of generated measurement points.

The conclusions can be supported by the results presented in Table VI, where the values for mse and $nrmse$ are given along with the ratio between the number of actually made measurements and the number of validation measurements. The model quality is, as expected, lower for the higher dimensional cases. In these cases, the ratio between the measurements and the validation dataset is very low. The results could be improved if the termination criteria were different, i.e., defined to optimize one of the prediction criteria.

VIII. CONCLUSION

This paper has presented the DoE based on supervised, hierarchical clustering, and fuzzy model structure. The benefits of the proposed approach are an ability to model the nonlinear system with a small number of local models and a small number of measurements in the input domain. The demonstration example shows that the proposed method, which combines the advantages of the evolving fuzzy model with fuzzy clustering and the maximin experimental plan, is able to identify and model the behaviors that are very complex and highly nonlinear.

REFERENCES

[1] R. Precup, M. Tomescu, and S. Preitl, "Lorenz system stabilization using fuzzy controllers," *Int. J. Comput., Commun. Control*, vol. 2, no. 3, pp. 279–287, 2007.

[2] S. S. Auephanwiriyakul and J. Keller, "Analysis and efficient implementation of a linguistic fuzzy c-means," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 5, pp. 563–582, Oct. 2002.

[3] G. Taguchi, "Quality engineering (Taguchi methods) for the development of electronic circuit technology," *IEEE Trans. Rel.*, vol. 44, no. 2, pp. 225–229, Jun. 1995.

[4] D. C. Montgomery, *Design and Analysis of Experiments*. New York, NY, USA: Wiley, 2012.

[5] B. Hartmann, J. Moll, O. Nelles, and C.-P. Fritzen, "Hierarchical local model trees for design of experiments in the framework of ultrasonic structural health monitoring," in *Proc. IEEE Int. Conf. Control Appl.*, 2011, pp. 1163–1170.

[6] T. W. Simpson, J. Poplinski, P. N. Koch, and J. K. Allen, "Metamodels for computer-based engineering design: Survey and recommendations," *Eng. Comput.*, vol. 17, no. 2, pp. 129–150, 2001.

[7] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. New York, NY, USA: Springer Verlag, 2003.

[8] D. A. Cohn, "Neural network exploration using optimal experiment design," *Neural Netw.*, vol. 9, no. 6, pp. 1071–1083, 1996.

[9] C. Hametner, M. Stadlbauer, M. Deregnaucourt, S. Jakubek, and T. Winsel, "Optimal experiment design based on local model networks and multilayer perceptron networks," *Eng. Appl. Artif. Intell.*, vol. 26, pp. 251–261, 2013.

[10] D. A. Cohn *et al.*, "Minimizing statistical bias with queries," *Adv. Neural Inform. Process. Syst.*, vol. 1552, no. 124, pp. 417–423, 1997.

[11] B. Hartmann, T. Ebert, and O. Nelles, "Model-based design of experiments based on local model networks for nonlinear processes with low noise levels," in *Proc. Amer. Control Conf.*, 2011, pp. 5306–5311.

[12] B. Hartmann, O. Banfer, O. Nelles, A. Sodja, L. Teslić, and I. Škrjanc, "Supervised hierarchical clustering in fuzzy model identification," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 6, pp. 1163–1176, Dec. 2011.

[13] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modelling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, pp. 116–132, Jan./Feb. 1985.

[14] H. G. Ying, "Necessary conditions for some typical fuzzy systems as universal approximators," *Automatica*, vol. 33, pp. 1333–1338, 1997.

[15] L.-X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 807–814, Sep. 1992.

[16] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proc. IEEE 17th Symp. Decision Control Including Adaptive Processes*, 1978, vol. 17, pp. 761–766

[17] H. B. N. O. Teslić, Luka, and I. Škrjanc, "Nonlinear system identification by Gustafson–Kessel fuzzy clustering and supervised local model network learning for the drug absorption spectra process," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 1163–1176, Dec. 2011.

[18] M. E. Johnson, L. M. Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *J. Stat. Plan. Inference*, vol. 26, no. 2, pp. 131–148, 1990.

[19] P. Goos and B. Jones, *Optimal Design of Experiments: A Case Study Approach*. New York, NY, USA: Wiley, 2011.

[20] M. Stadlbauer, C. Hametner, S. Jakubek, and T. Winsel, "Analytic multilayer perceptron based experiment design for nonlinear systems," in *Proc. 8th IFAC World Congr.*, 2011, pp. 4332–4337.

[21] J. Sokolić and I. Škrjanc, "Design of experiments based on fuzzy model identification with supervised hierarchical," in *Proc. 20th East West Zittau Fuzzy Colloq.*, 2013, pp. 198–205.

[22] M. Melouna, D. Burkoňová, T. Syrovy, and A. Vrána, "Thermodynamic dissociation constants of silychristin, silybin, silydianin and mycophenolate by the regression analysis of spectrophotometric data," *Anal. Chim. Acta*, no. 486, pp. 125–141, 2003.

[23] R. Babuska, *Fuzzy Modeling for Control*. Norwell, MA, USA: Kluwer, 1998.

[24] J. Jang, "ANFIS: Adaptive-network-based fuzzy inference systems," *IEEE Trans. Syst., Man Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.

[25] O. Nelles, *Nonlinear System Identification*. Berlin, Germany, Springer, 2001.

[26] D. Gorissen, L. De Tommasi, W. Hendrickx, J. Croon, and T. Dhaene, "RF circuit block modeling via kriging surrogates," in *Proc. Microwaves, Radar Wireless Commun.*, 2008, pp. 1–4.

[27] J. Sokolić, "Fuzzy model based design of experiments," Bachelor's work under supervision of I. Škrjanc, Univ. Ljubljana, Ljubljana, Slovenia, pp. 54-60, 2013.

Igor Škrjanc received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Faculty of Electrical and Computer Engineering, University of Ljubljana, Ljubljana, Slovenia, in 1988, 1991, and 1996, respectively.

He is currently a Professor of automatic control with the Faculty of Electrical Engineering, University of Ljubljana, and the Head of the research program Modeling, Simulation and Control. His main research interests include intelligent, predictive control systems, and autonomous mobile systems.

Dr. Škrjanc received the Highest Research Award of the University of Ljubljana, Faculty of Electrical Engineering, in 2007, the Highest Award of the Republic of Slovenia for Scientific and Research Achievements in 2008, and the Zois Award for outstanding research results in the field of intelligent control. He also received the Humboldt Research Fellowship for Experienced Researchers between 2009 and 2011. In 2012, he was a member of the team that won the competition organized by IEEE Computational Society known as Learning from the data in the frame at the IEEE World Congress on Computational Intelligence 2012, Brisbane, Australia: Solving the sales prediction problem with fuzzy evolving methods. In 2013, he received the Best Paper Award at the IEEE International Conference on Cybernetics, Lausanne, Switzerland. He is also an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS; the IEEE TRANSACTIONS ON FUZZY SYSTEMS, and *Evolving Systems*.