Detecting ambiguities in regression problems using TSK models

ARUP KUMAR NANDI

Central Mechanical Engineering Research Institute, M.G. Avenue, Durgapur-713209, WB, India

nandiarup@yahoo.com

FRANK KLAWONN

Department of Computer Science, University of Applied Sciences BS/WF, Salzdahlumer Str. 46/48, 38302, Wolfenbuettel, Germany

f.klawonn@fh-wolfenbuettel.de

Abstract: Regression refers to the problem of approximating measured data that are assumed to be produced by an underlying, possibly noisy function. However, in real applications the assumption that the data represent samples from one function is sometimes wrong. For instance, in process control different strategies might be used to achieve the same goal. Any regression model, trying to fit such data as good as possible, must fail, since it can only find an intermediate compromise between the different strategies by which the data were produced. To tackle this problem, an approach is proposed here to detect ambiguities in regression problems by selecting a subset of data from the total data set using TSK models, which work in parallel by sharing the data with each other in every step. The proposed approach is verified with artificial data, and finally utilised to real data of grinding, a manufacturing process used to generate smooth surfaces on work pieces.

Keywords: Ambiguities; Linear regression; TSK fuzzy model

1. Introduction

It is obvious that having as much data as possible is beneficial for regression problems, as the data are used to construct and validate a model. Moreover, the information content of the data set used to construct a model should be correct and maximized in order to minimize the erroneousity of the output results obtained by the developed model. However, it is very difficult to identify the suitable data in a large data set of a physical process beforehand. In general, these data are taken based on observation or measured using suitable measuring instruments during process operation. Often it is not possible to collect good data using real experimentation, because of the improper experimental set up as well as instrumental errors.

Another difficulty that occurs in regression problems is that the data do not represent (noisy) measurements of a single function, but (noisy) measurements from more than one function. This can happen for different reasons. When the data are taken from observing human operators controlling a process, there might be different strategies used by the same or different operators to achieve the same goal. As a simple example, consider steering a vehicle including obstacle avoidance. An obstacle can be avoided by either steering to the right or to the left. Data from both cases might be present in the data set. Any regression function trying to minimize the mean (squared) error can only interpolate between the two strategies which leads to bumping straight into the obstacle as illustrated in Fig. 1.

Another reason for the presence of such ambiguities in data sets might occur in the case, when there are attributes that contain important information that were not observed or not observable. Therefore, for the same values of the observed attributes we might have completely different values for the dependent variable due to different values in the non-observed attributes.

Another example of ambiguity lies in the fact of selecting the optimal values of the machining parameters in manufacturing processes to achieve a desired output such as surface finish. In a grinding process (a manufacturing process to generate a smooth surface on a work piece), the selection of optimal values of wheel speed to achieve a desired surface finish with different values of wheel dress lead is ambiguous in the operator's point of view for a given work speed and feed rate. Practically, different operators have used their own strategy to consider the values of parameters to perform a particular task.

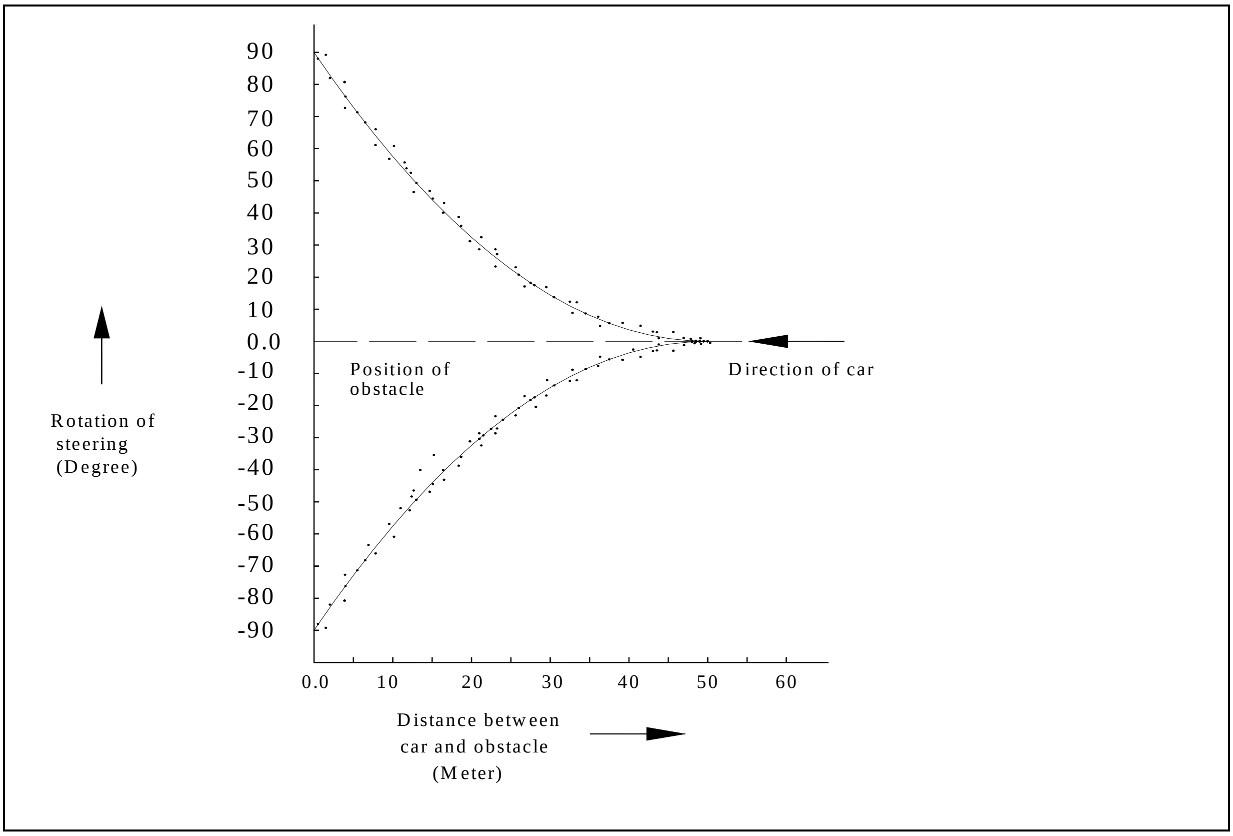


Fig. 1. Artificial data of steering a vehicle to avoid an obstacle

Again, there is also the possibility of existing incorrect data due to the inherent human error, which cannot be avoided. Therefore, the existence of imperfect data in the whole data set as well as data collected from various sources; based on which regression analysis is carried out to develop a model, affect the model's performance. This is obvious because every data point involved in a particular data set affects the parameters of the regression function to be computed.

Thus, there exists a problem in selecting data for determining the regression function(s) of a system from the whole data set, especially when data are obtained from different sources, in order to achieve the best fit. The main objective of this research work is to detect and identify the subset(s) of data form the entire data set (universe), that relate to different possible strategies or models. Therefore, it is a problem of using the application of parallel or multiple models for the detection of ambiguities. The idea of multiple models is especially used for the on-line diagnosis of industrial applications (fault detection) [26] or for some special types of adaptive controllers [17].

To detect these ambiguities in selecting the data for a regression problem, in this work, a simple and effective approach is proposed using TSK fuzzy models. The idea of the proposed approach is quite similar to the working procedure of (fuzzy) switching regression method [11, 12, 16, 24]. In switching regression the data is partitioned among a set of regression models according to how a given model can describe the data. In [6], Shen et al. described an algorithm, called FCWRM (Fuzzy C weighted regression model) to detect outliers in fuzzy switching regression models. However, in this proposed approach, the operating regimes of the local linear models are defined by the antecedent membership functions of the fuzzy models.

The performance of a TSK fuzzy model depends on the values of the coefficients of the functions, which characterize the consequents of the fuzzy rules. Each rule of a TSK model is considered as a regression function. A linear regression analysis is carried out to determine the coefficients of the output functions of each rule of the TSK models based on the squared-error cost [10], which is commonly referred to as the regression error. The TSK fuzzy models are considered to work simultaneously by sharing the data with each other in every stage. It is important to mention that the set of data selected for each TSK model depends on the accuracy level for each TSK model. The characteristics of the TSK models are analysed for different sub-regions in the input space by comparing the values of function coefficients of the rule consequents of each TSK model. The proposed approach is explained with the artificial data of the problem in steering a vehicle to avoid an obstacle using two TSK models. We also discuss an example of a real data set representing the problem of selecting the wheel speed to produce surface finish on a work piece with different values of wheel dress lead in a grinding process, a manufacturing process used to generate a smooth surface on a work piece. However, the number of TSK models required to fit to the ambiguous data of a physical process is an important aspect.

The rest of the paper is arranged as follows: Section 2 describes the TSK fuzzy model. The proposed approach to identify the inherent ambiguities in a data set for a regression analysis is discussed in Section 3. Section 4 illustrates the application of the proposed approach using real data of a grinding process. Some concluding remarks are given in Section 5.

2. TSK fuzzy model

The TSK fuzzy model is based on the fuzzy rule-based system, which is proposed by Takagi, Sugeno and Kang [5, 4]. The overall output of the TSK fuzzy model can be obtained for the input tuple (x_1, x_2, …., x_n) using the following empirical expression.

$$Y = \frac{\sum_{r=1}^{R}\left(\prod_{v=1}^{n}\mu_v^r(x_v)\right)\sum_{j=1}^{K} a_j^r f_j^r(x_1,\ldots, x_n)}{\sum_{r=1}^{R}\left(\prod_{v=1}^{n}\mu_v^r(x_1,\ldots, x_n)\right)} \tag{1}$$

where n is the number of input variables that occur in the rule premise, R is the number of rules in the rule base. $\prod_{v=1}^{n}\mu_v^r(x_1,\ldots, x_n)=\eta_r$ is the firing degree of the r^{th} rule. $\prod$ is the product representing a conjunction. Instead of the product, other t-norms can also be used.

$\sum_{j=1}^{K} a_j^r f_j^r(x_1,\ldots, x_n)$ is the output function of the r^{th} rule and a_j are the function coefficients of the corresponding rule consequent. K is the number of parameters (coefficients) associated to a regression function.

The TSK-fuzzy model has the following form of fuzzy rules:

If x_1 is A^r_1 and x_2 is A^r_2 and…and x_n is A^r_n, then $y^r = f^r(x_1,\ldots, x_n)$.

where $A^r_1, \ldots, A^r_n$ are fuzzy subsets of the input variables $x_1, \ldots, x_n$, respectively. The output function of each rule is considered as a function in the form

$$y^r = \sum_{j=1}^{K} a_j^r f_j^r(x_1,\ldots, x_n) \tag{2}$$

The performance of a TSK fuzzy model mainly depends on the optimal values of the output function coefficients of the rules and also on the choice of the type of fuzzy sets. A linear regression analysis is conducted based on available data to determine the coefficients.

It is obvious that Equation (1) is a linear function in the parameters a_j^T, so that they can be determined by a standard least square technique [20, 22] leading to a system of linear equations, when we assume that the rules and the fuzzy sets are assumed to be fixed.

3. Splitting the data set in order to detect ambiguities in regression problems using TSK models

In our proposed approach for the identification of ambiguities in data for regression problems (RP), a number of predefined TSK fuzzy models are considered. When we try to fit the TSK models to the data set, these TSK models are allowed to work simultaneously in each fitting step by sharing or compete for the imperfect data of one model by the other in an iteration procedure. This idea of data competing or sharing is implemented following the concept of alternating optimization in clustering. Of course, since no analytical solution exists for such a complex optimisation problem, the algorithm might get stuck in local minima. It is therefore recommended to run the algorithm more than once with different initialisations. The algorithm resembles a crisp clustering technique - the assignment of the data to the different TSK models - where parameters of the cluster prototypes are the parameters of the corresponding TSK model. Later on, additional global parameters will be considered that define the fuzzy partitions of the input space. These parameters will be optimised by an evolutionary algorithm.

In the following, we explain our methodology in more detail using two and three TSK models in subsection 3.1. In subsection 3.1, we have tested the proposed methodology by using one-dimensional artificial data of steering a vehicle to avoid an obstacle as illustrated in Fig. 1. Subsection 3.1 will also explain, how the fuzzy sets can be optimised simultaneously.

3.1. Using two TSK models

The data set is partitioned into two subsets, one for each TSK model. In the first step the data set is partitioned randomly into two subsets of approximately the same size. For each TSK model the coefficients are adapted to its assigned data subset by standard regression as it is described in Section 2. Then we compute the (absolute) errors of both TSK models for all data. Based on these errors, the data set is again partitioned into two subsets. A data point is assigned to that TSK model which yields the smaller error for this data point. Based on this new partition of the data set, we can re-compute the coefficients of the two TSK models, partition the data set again and continue this procedure until the process will finally converge. Note that convergence is always guaranteed. When we re-assign the data to the TSK models, the overall error can only be reduced (or will remain constant in the worst case), since we only assign a data point to the other TSK model, if the TSK model it has been assigned to before yields a larger error. When we re-compute the coefficients, the linear regression approach guarantees the optimal solution, when we fix the current assignment of the data to the two TSK models. Therefore, this step will also reduce the overall error. This means in each alternating optimization step, only reductions of errors can take place, so that convergence is always guaranteed. Fig. 2 illustrates the scheme of this algorithm using two TSK models.

It should be mentioned that the number of data assigned to each of the two different TSK models might not be equal. This iteration process is allowed to continue until a satisfactory result is obtained. The termination criteria of this algorithm is imposed such that, when there is no data found for TSK I which shows a higher error than that for TSK II. After certain iterations, the total data set will be divided into two subsets, one for TSK model I and the other subset for TSK model II.

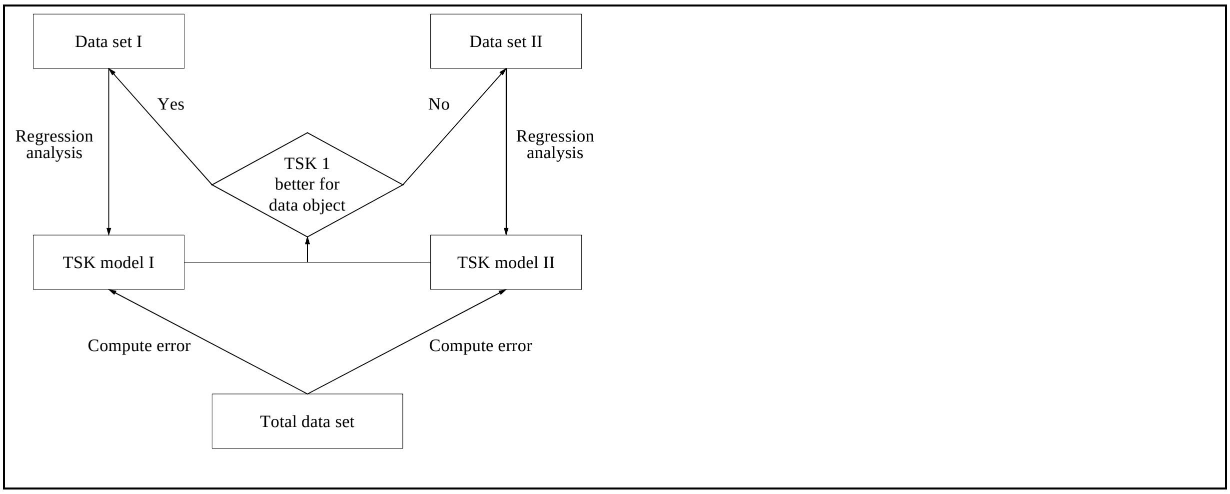


Fig. 2. Identification data set in order to detect ambiguities in RP using two TSK models

Now we present the iterative algorithm mathematically.

Algorithm: Identification data set in order to detect ambiguities in RP using two TSK models

1. Let $D = \{d_1, d_2, \ldots\ldots, d_N\}$ denotes the total dataset containing N input-output tuples.
2. Initialize the number of TSK models (M_i represents the ith TSK model, where i=1,2,…I, and $I \geq 2$.
3. Partition the total dataset into I number of data subsets (D_i^t) of (almost) equal sizes.
4. Determine coefficients of rule consequent functions $[C_{rk}]_i$ of each TSK models (M_i) using equation (5), where r=1,2,….,R; k=1,2,…,K, based on the data subset (D_i^t).
5. Compute the error values of the different TSK models (M_i) for each data sample,

$$|e_i|_{z=1,2,\ldots\ldots,N} = \left|Y_{M_{i,z}} - y_z\right|$$

6. Based on these error values, re-allocate the data samples in the different subsets. That is,

$$d_z \rightarrow D_i^{t+1}, \ \ if \ \ |e_i|_z < |e_p|_{z,p=1,2,\ldots.,I;p\neq i}$$

7. Test the convergence $D_i^t \equiv D_i^{t+1}$, if YES, stop algorithm
8. Otherwise $t \leftarrow t+1$, go to step 4.

Fig. 1 shows the artificial data with random noise of steering a car to avoid an obstacle. In this problem, the driver has two options to steer the car either to the left side (negative values of the rotation of steering) or to the right side (positive values of the rotation of steering) to avoid the collision with the obstacle, which lies in front of the car. The data points show the angle of rotation of the steering wheel (y) with respect to the distance between the car and the obstacle (x). Fig. 1 describes the problem; such that when the car is 50 meters away from the obstacle the rotation of the steering wheel is still zero, whereas a rotation of 90 degrees is chosen just before the collision of the car with the obstacle. To verify the performance of the proposed approach, the artificially created data that are shown in Fig. 1, are considered as the total data set. In this problem the shape of the membership function distributions of the input variable i.e., the distance between the car and the obstacle is considered as trapezoidal, whereas the angle or rotation of the steering is the output variable. To keep the simplicity of the problem the structure of the output function of each fuzzy rule is of the same form (but allowing different coefficients). Furthermore, the same configuration of the rule consequent functions is considered for both the TSK fuzzy models. In addition to this, no such on-line optimisation of fuzzy partitions of the input space is considered during the construction for the TSK models. The structure of the output function of each fuzzy rule of the TSK models is chosen based on some experimental test cases, as defined in the following form:

$$y = C_1 x + C_2 x^{2..0} \tag{3}$$

where C_1 and C_2 are the function coefficients. After the fifth iteration the total data set is perfectly divided into two subsets. The membership function distribution of the data points for both the TSK models (TSK model 1 and TSK model II) are shown in Fig 3. The values of coefficients of the rule consequent functions of the TSK models are also depicted in Table 1. It is important to mention here that the number of iterations required to obtain the best result is based on the structure of the output functions of the rules to fit the data. It has been observed that the values of the coefficients of the TSK model that is constructed based on the entire data samples as illustrated in Fig. 1, give erroneous results (not shown here).

Sometimes, it may occur that the parameters of the local linear models (rule consequent functions) of different fuzzy systems converge to the same result (local optima of the complex optimization algorithm behind the proposed iterative algorithm). This problem can be solved using predefined constraints on the parameters [31].

In some cases, it might happen that the chosen number of fuzzy sets, i.e. regions for the TSK model, is too small and in one or more of the regions, a single TSK model might not be able to fit the data. Not, because the data are ambiguous, but simply because the model is not flexible enough. We therefore recommend to further investigate the regions where more than one TSK model is valid. The input data in such a region should be inspected using visualisation techniques like multi-dimensional scaling or clustering. When it turns out that the data for the different TSK models are well separated or belong to different clusters, this is an indicator for the case that the two TSK models do not represent ambiguity but a more complex model in that region.

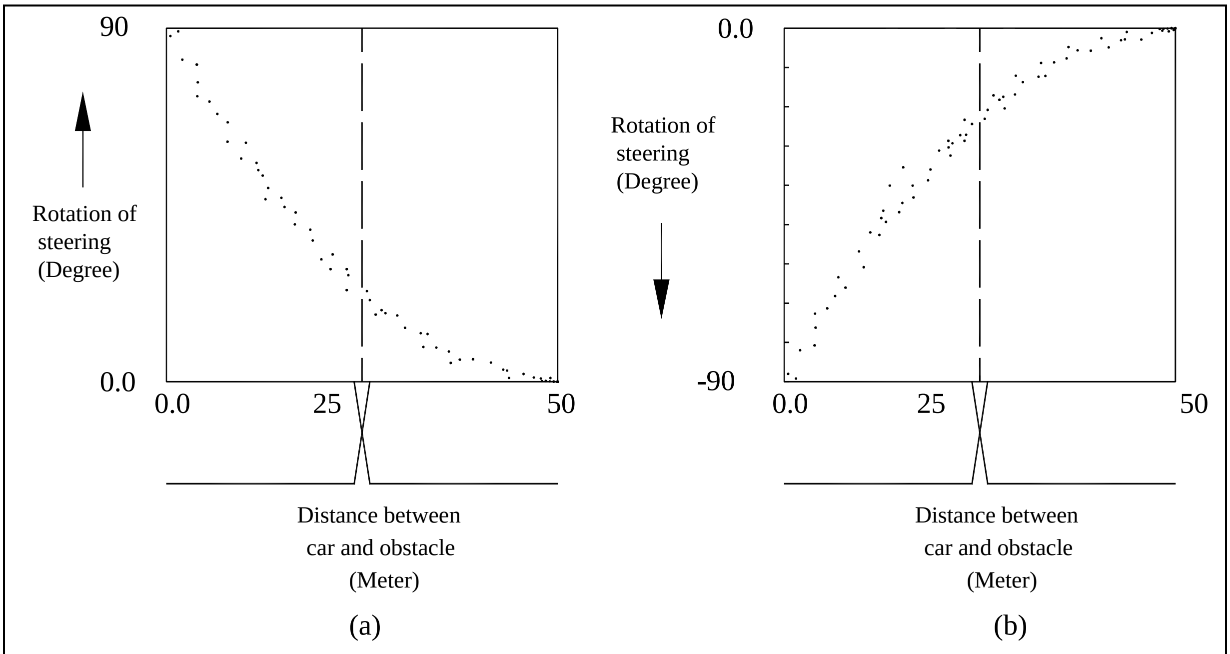


Fig. 3. Membership function distributions of input variable (x) (a) TSK model I, (b) TSK model II

Table 1

Coefficient values of rule consequents of TSK models

(a) TSK model I

No of rules	C_1	C_2
Rule 1	11.12530	-0.46605
Rule 2	1.57539	-0.03573

(b) TSK model II

No of rules	C_1	C_2
Rule 1	-11.12510	0.46604
Rule 2	-1.23855	0.02570

The method of detecting the ambiguities in regression problems using two TSK models as described above may be utilized by adopting three or more TSK-type fuzzy models. Here, we have assumed predefined fuzzy partitions for the fuzzy rules. Only the coefficients of the rule consequents are optimized and, of course, the assignment of the data to the two TSK models. Incorporating the optimization of the fuzzy partitions will be the topic for the next paragraph. There are several methodologies adopted by various researchers in order to optimize the membership function distributions such as neural networks (NNs) [27], genetic algorithm (GA) [23, 30], combined GA and NN [2], evolutionary algorithms [25], Tabu search [29], an extended Kalman filter algorithm [7], simulated annealing [28], etc. In the following subsection, we extend our proposed methodology to simultaneously optimising the membership functions using an evolutionary algorithm (EA).

In this algorithm as described in Fig. 4, three TSK models are used in order to detect the ambiguities in the regression problem. As in the previous approach of using two TSK models, here these three TSK models are allowed to work simultaneously in each step by sharing or compete for the imperfect data with each other in the next step. In the first step, the whole data set is divided randomly into three subsets of approximately the same size. For each TSK model, the rule coefficients are determined using the respective assigned data subset by standard regression as it is described in Section 2. Simultaneously, the membership function distributions of the input variables of the TSK models are optimized using an evolutionary algorithm. After that, the (absolute) errors of the three TSK models for all data points are computed and the whole data set is re-partitioned into three subsets (one for each TSK model) again based on the errors of the corresponding models. A data point is assigned to that TSK model, which yields the smallest error for this data point. This iteration procedure is allowed to continue until the process will converge. The termination criterion (equilibrium state condition) of this iteration process is kept the same as in the previous approach using two TSK models, that is when data points no longer migrate from one TSK model to another.

In Section 4, we have implemented this technique using real data of a grinding process in order to determine the optimal values of wheel speed to achieve a desired surface finish with different values of wheel dress lead for a given work speed and feed rate.

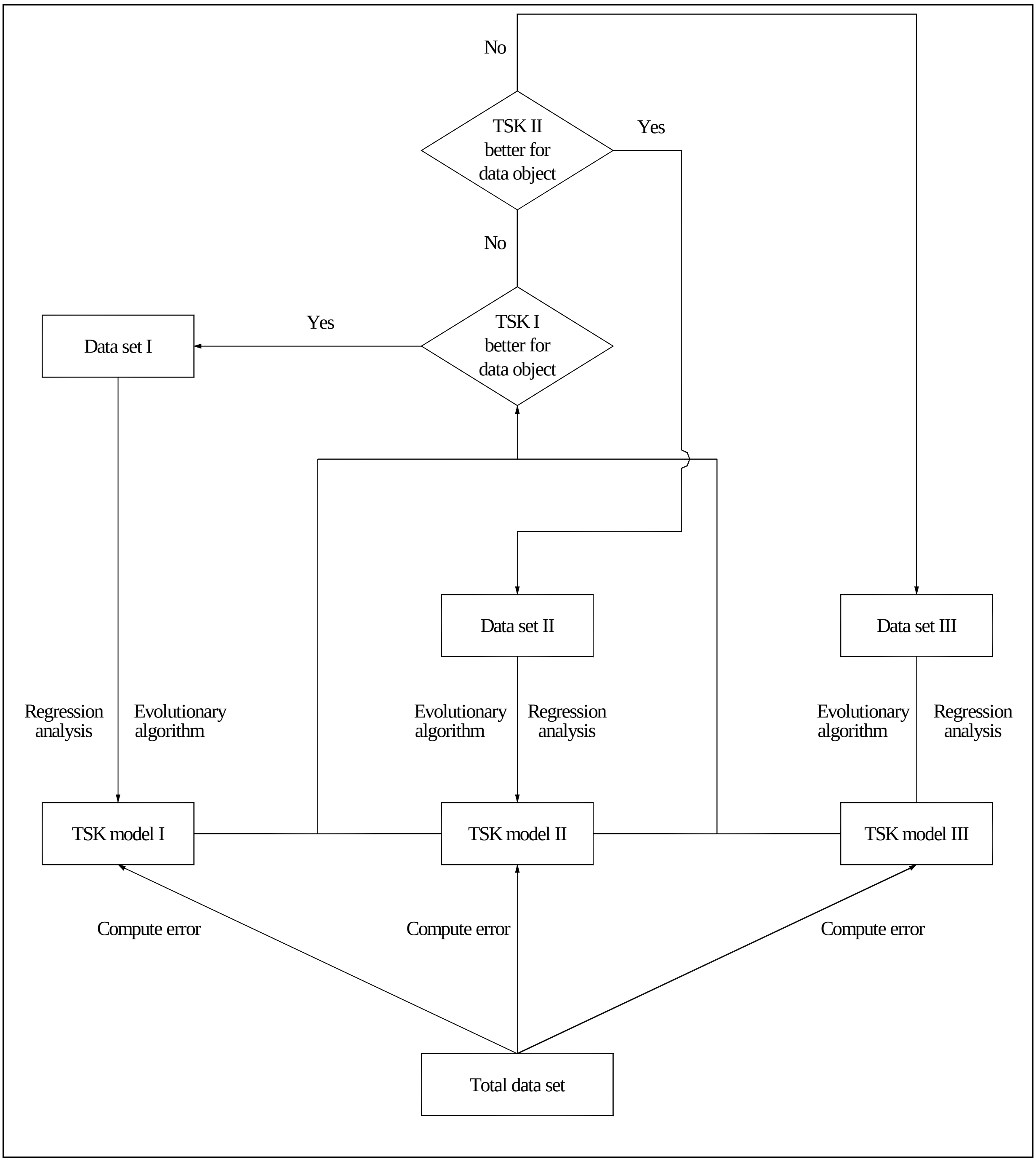


Fig. 4. Identification data set in order to detect ambiguities in RP using three TSK models

4. Application using real data in a grinding process

Grinding is a well-known manufacturing process especially used for machining of hard material [21]. It is a complex manufacturing process that is used to achieve a desired geometry and smooth surface of a work piece using an abrasive tool, called wheel. Surface finish in grinding processes is mainly affected based on the optimal values of the cutting parameters namely, wheel speed for a given feed rate and work piece speed with the wheel dress lead. There are many approaches proposed by various authors where intelligent systems including fuzzy logic [9, 1, 3], neural networks [18, 8, 15], fuzzy basis function networks (FBFN) [14,19], genetic fuzzy (GA-Fuzzy) approaches [13], etc were extensively used for grinding processes. However, in those works, the main objectives of the authors were to design a suitable single model based on the whole available data samples irrespective of finding how many models may fit to these data. Sometimes it can be observed that such a physical process follows different functions over the same interval(s) of the process variable(s). In this case, it does not make any sense to determine a single optimum model using the data samples that follow different functions/models, since the outputs of those (optimal) models show erroneous results. Thus our primary objective is to analyze the presence of such ambiguities in the cylindrical plunge grinding process using the proposed approach of detecting ambiguities in regression problems using TSK fuzzy models.

For this application 300 real data of wheel speed for surface roughness with different wheel dress lead in a cylindrical plunge grinding process are considered. Fig. 5 illustrates the values of wheel speed for different cases (surface finish and wheel dress lead) taken from various operators' points of view. The wheel speed is measured in revolutions per minute (rev/min), whereas surface finish and wheel dress lead are taken in micro-meter (micron) and millimetre (mm), respectively. The data of wheel speed for different surface roughness values are taken for the fixed values of feed rate and work piece speed of 0.5 mm/min (millimetre per minute) and 112 rev/min, respectively.

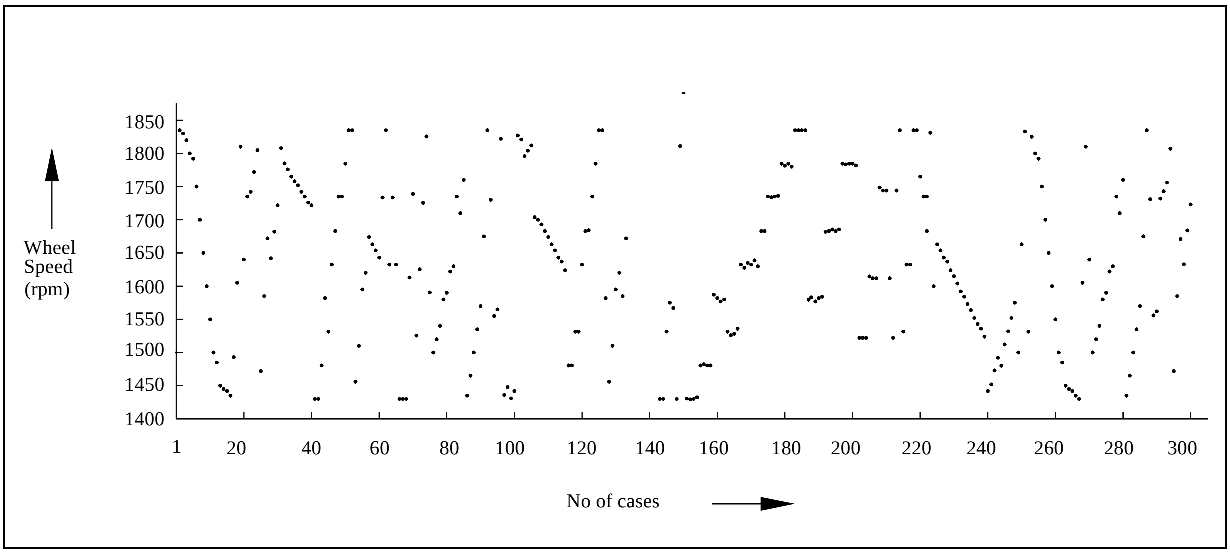


Fig. 5. 300 real data of wheel speed in a grinding process used in RP

We have partitioned the entire range of each of the input variables (wheel dress lead and surface roughness) by two fuzzy sub-sets of semi-trapezoidal shape. Thus the whole input space is divided into four regions as shown in Fig. 6. Since each of the input variables has two fuzzy sets within its range, there is a maximum of $2 \times 2 = 4$ rules present in the rule base of each TSK fuzzy model and each rule corresponds to an input-region of Fig. 6. In this problem, the structure of the output function of each fuzzy rule of the TSK models is chosen based on some experimental test results, in the following form:

$$f(p_1, p_2) = C_3 \times p_1 + C_4 \times p_2 + C_5 \times p_1^2 \times p_2^2 \quad (4)$$

where p_1 and p_2 stand for the input variables, surface finish and wheel dress lead, respectively. C_3, C_4 and C_5 denote the function coefficients.

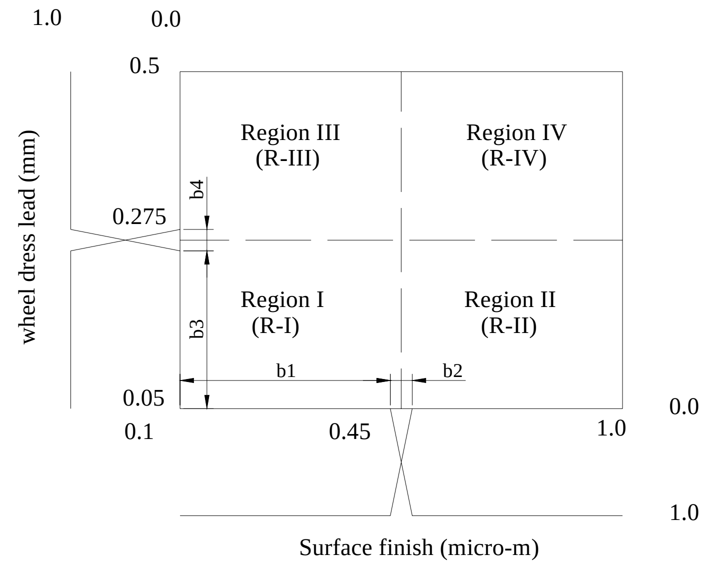


Fig. 6. Input space with semi-trapezoidal membership function distribution

However, since an arbitrary partition of the input-space is not at all an optimal one, in this work an evolutionary algorithm (EA) is adopted to determine the optimal partitions/membership function distributions of the input variables. For this practical application, we have utilized the algorithm of detecting ambiguities in regression problems using two TSK models (Approach 1), three TSK models (Approach 2) and using four TSK models (Approach 3).

EAs are global search techniques derived from Darwin's theory of evolution by natural selection. An EA iteratively updates a population of potential solutions, which are often encoded in structures called chromosomes. During each iteration step, called a generation, the EA evaluates solutions and generates offspring based on the fitness of each solution in the task environment. Substructures, or genes, of the solutions are then modified through genetic operators such as mutation and crossover. The idea is that structures that are associated with good solutions can be mutated or combined to form even better solutions in subsequent generations. In order to optimise the membership function distributions of input variables, the string length of chromosomes in EA is taken as 80 bit long, keeping 20 bit for each of the four parameters (b1, b2, b3 and b4, as shown in Fig. 6). The performance of an EA mainly depends on the optimal settings of the parameters namely population size, crossover probability and mutation probability. However, the parameters of the EA in all three approaches are kept fixed to facilitate the better comparison among them. We have kept the values of different parameters of the EA as follows: Population size=100; Number of generations=50; Crossover probability=0.82, Mutation probability=0.012. The fitness value of each solution of the population is taken as the root mean square (RMS) error value of the corresponding data samples. The performances of the TSK models obtained using different approaches are discussed in the following sub-section.

4.1. Results and discussion

As discussed earlier in subsection 4.1, the number of iterations required to obtain the best selection of a data subset depends on the rule consequent functions that we have chosen for the TSK models. Furthermore, the number of data selected for each model is also varying with the different forms of the rule output functions. Based on the chosen form of the rule consequent function (see Equation (4)), the number of iteration steps required for different approaches are listed in Table 2.

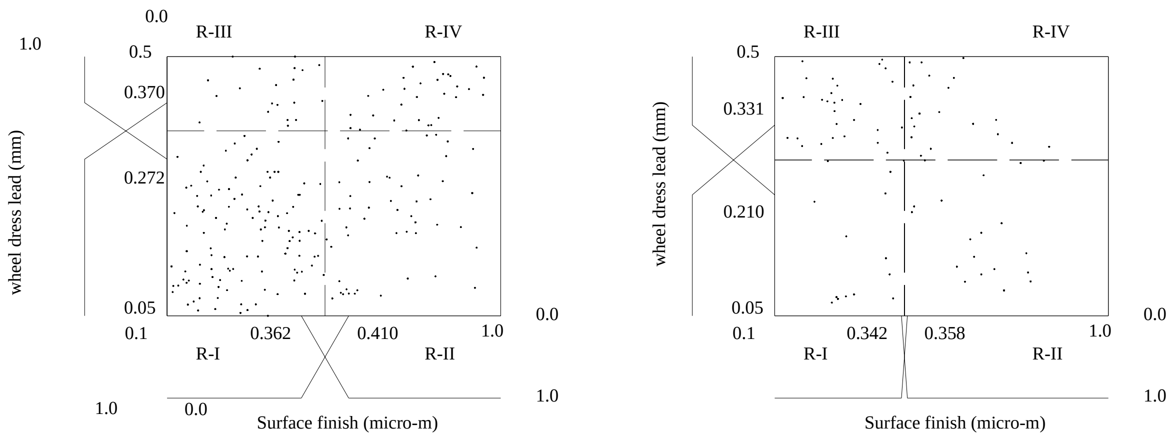
Table 2

Number of iteration steps required and data sample selected by TSK models in different approaches

Approach	Number of iteration steps required	Number of data sample selected by TSK model			
		TSK model I (Data set I)	TSK model II (Data set II)	TSK model III (Data set III)	TSK model IV
Approach 1	20	226	74	———	———
Approach 2	35	158	56	86	———
Approach 3	48	56	158	86	NIL

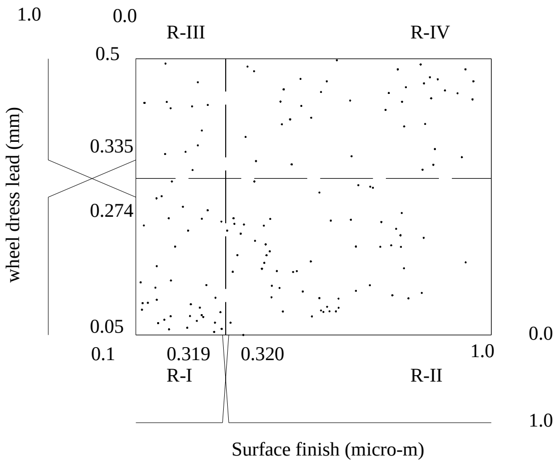
In Table 2, it can be seen that out of the 300 data, 226 data were selected for TSK model I, and the remaining 74 data for TSK model II in Approach 1, whereas in Approach 2, TSK model I, TSK model II and TSK model III select 158, 56 and 86 data samples, respectively. Fig. 7 describes the data points selected by TSK model I and TSK model II using Approach 1. Whereas, the data samples assigned for different TSK models (158 for TSK model I, 56 for TSK model II and 86 for TSK model III) in Approach 2 are shown in Fig. 8. In Approach 3, we can observe that no data are assigned to TSK model IV, whereas the data assigned to other TSK models (TSK model I, TSK model II and TSK model III) are identical to those obtained by the TSK models (TSK model II, TSK model I and TSK model III) in Approach 2. Fig. 9 illustrates the number of data selected for different TSK models at each iteration steps of Approach 3. At the 13th iteration step, one of the TSK model (TSK model IV) looses all its data. From this phenomenon, we may conclude that the data samples from the grinding process we have considered here are actually generated by three different functions. However, in Approach 3, the algorithm needs more iteration steps (48) to reach convergence than that (35) required in Approach 2.

The optimized membership function distributions of the input variables for the different TSK models obtained using the EA in Approach 1 and Approach 2 are shown in Fig. 7 and Fig. 8, respectively.

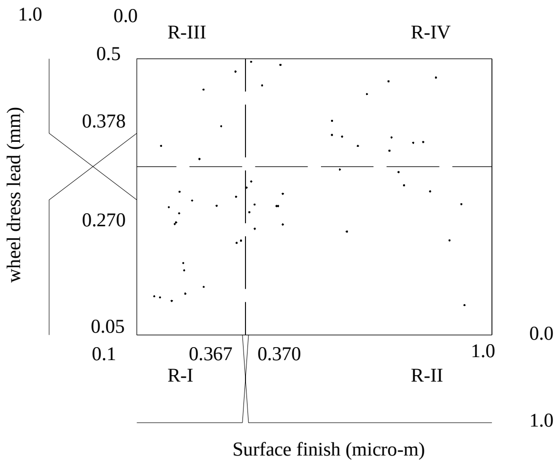


(a) (b)

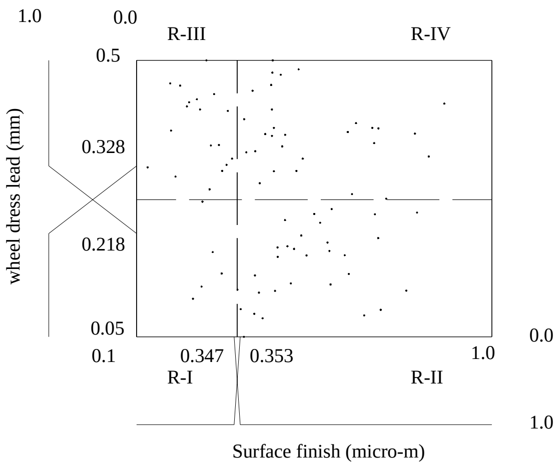
Fig. 7. Optimized membership function distributions of input variables using Approach 1 (a) TSK model I, (b) TSK model II



(a)



(b)



(c)

Fig. 8. Optimized membership function distributions of input variables using Approach 2 (a) TSK model I, (b) TSK model II and (c) TSK model III

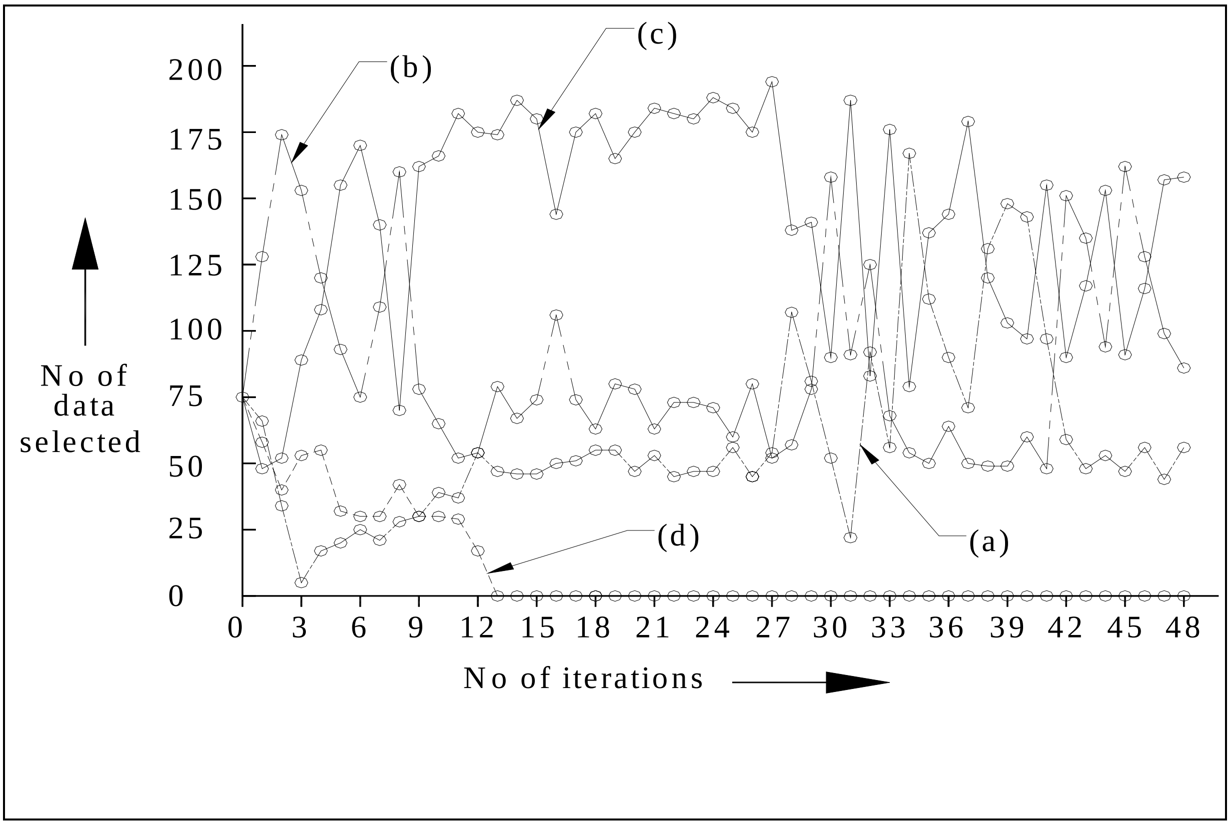


Fig. 9. Number of data selected in different iteration steps of Approach 4 (a) TSK Model I, (b) TSK Model II, (c) TSK Model III and (d) TSK Model IV

The performance of the two TSK models whose rule consequent functions coefficients are determined based on Approach 1 are enlisted in Table 3. The entries shown in Table 3 are the root mean square (RMS) errors of the output values of the TSK models corresponding to the data sets (Data set I and Data set II). β stands for the RMS error value of the outputs for all the data samples obtained by the single TSK models whose rule consequent function coefficients are determined based on all (300) data samples.

Table 3
Results of the RMS error values of the TSK models using Approach 1

Data set	TSK I	TSK II
Data set I	0.914 β	2.179 β
Data set II	1.763 β	0.434 β
Total data set	1.185β	1.901β

The RMS error value of TSK II based on data set II (84 number of data samples) is significantly smaller than β, while that of TSK model I which is constructed based on data set

I (226 data samples) is not significantly smaller. From this point of view, we might suspect that data set I contains data samples that originate from more than one model. It can also be observed that the RMS error of TSK model I whose rule consequent functions are determined based on the data set I is less for data set I compared to the total data set as well as data set II. The same kind of result is also found in case of TSK model II for data set II compared to the total data set and data set I.

Table 4

Results of the RMS error values of the TSK models using Approach 2

Data set	TSK I	TSK II	TSK III
Data set I	0.509 β	2.127 β	29.102 β
Data set II	16.789 β	0.2655 β	37.293 β
Data set III	2.914 β	1.896 β	0.412 β
Total data set	15.143β	23.076β	2.267β

Table 4 shows the results of different TSK models those are obtained using Approach 2. In Table 4, it has been noticed that the RMS error values of all models (TSK model I, TSK model II and TSK model III) obtained based on Approach 2 are significantly smaller than β. In this case, the same phenomena as observed in Table 3 is that the RMS error value of one TSK model whose rule consequent functions are determined based on the corresponding data set is less for the same data set compared to those are obtained for other data sets as well as the total dataset.

After analysing the results of different approaches (Approach 1 and Approach 2) as illustrated in Table 3 and Table 4, it can be observed that the mean square errors of the TSK models that are constructed based on Approach 2 are smaller than that of the TSK models obtained using Approach 1. Furthermore, in Approach 3, the four models that we have considered initially are reduced to three models by assigning no data to the 4th model. The same phenomenon is also found in the case of running the algorithm using initially five or more models. In this case three models are allotted with data point whereas to the other models no data are assigned. This is another hint that the data samples we have considered here follow three different functions/models. However, this scenario is highly depended on the data samples as well as its erroneousity/noisiness. Thus, it is required to do more experimentation with larger data sets, preferably uniformly distributed over the span of input variables in order to obtain the different function(s) that are involved in the physical process.

The coefficients of the rule consequents for different TSK models obtained using Approach 1 and Approach 2 are enlisted in Table 5 and Table 6, respectively. The numerical values of the coefficients depend on the structure of the rule consequent function.

Table 5

Coefficient values of rule consequents of TSK models using Approach 1

(a) TSK model I

No of rules	C_3	C_4	C_5
Rule 1	73862	26971.5	-79949.2
Rule 2	22865.9	9762.1	-22293.8
Rule 3	-6699.36	-7059.17	20151.9
Rule 4	-4298.66	867.194	7465.13

(b) TSK model II

No of rules	C_3	C_4	C_5
Rule 1	151604	49094.1	-161824
Rule 2	1004410	-1782580	-1992290
Rule 3	-7544.39	-3637.72	14560.5
Rule 4	9911.3	10430.5	-17285.4

Table 6

Coefficient values of rule consequents of TSK models using Approach 2

(a) TSK model I

No of rules	C_3	C_4	C_5
Rule 1	11697.1	10431.8	-15373.6
Rule 2	8112.48	8005.27	-10084.7
Rule 3	264352	385449	-638181
Rule 4	71226.2	71556.9	-139675

(b) TSK model II

No of rules	C_3	C_4	C_5
Rule 1	-81180.5	-45923.5	128144
Rule 2	230727	117956	-329521
Rule 3	-92541.2	-69731	165846
Rule 4	-34436.5	-23388.8	61639.2

(c) TSK model III

No of rules	C_3	C_4	C_5
Rule 1	-4572.52	-12852.8	31825.2
Rule 2	16557.4	9154.34	-18398
Rule 3	-14947.5	-15201.1	35794.1
Rule 4	5191.01	4753.08	-6074.51

After analysing the results of different approaches, it can be stated that the proposed algorithm is able to detect ambiguities present in a mixture of data that originate from different functions or models. However, the choice of the initial number of models for the algorithm is an important issue. If the initial number of models is less than the actual number of models inherent in the data, the resulting models will not perform well. On the other hand, overestimating the number of models, the algorithm will need more iterations and more computation time. If this is acceptable, superfluous TSK models will automatically deleted by assigning no data to them as our experiments have shown.

5.2. Cross-validation of the proposed algorithm

In order to examine the performance of the proposed algorithm, we have considered 20 additional test cases (experimental data samples) that were not among the 300 training cases (thereby, the whole dataset consists of 320 cases). We have cross validated the results (RMSE value) of the TSK models obtained using Approach 1 and Approach 2 against Approach 0, based on these (20) test cases with the RMSE values based on the (300) training cases as shown in Table 7. In Approach 0, only one TSK model is constructed using all (300) training data as illustrated in Fig. 5. In this approach, the membership function distributions of the input variables of the TSK models are also optimized using an evolutionary algorithm with the same values of EA parameters used in other approaches as discussed in Section 5. For each sample the squared error is calculated with respect to the best result obtained among the TSK models of an approach and then we take the average of these values, and the square root of this average value is defined as the RMSE value. From Table 7, it can be seen that the RMSE values obtained using training data and test data in Approach I, are comparatively less than that obtained in Approach 0. On the other hand, the performances of the TSK models in Approach 2 are much better that of the TSK models obtained in both Approach 0 and Approach 1. The results of this cross-validation suggest that the model(s) obtained using the proposed algorithm does not only fit the training data rather it originally implies the underlying model(s) of the process.

The proposed approach extracts the information on ambiguities involved in the jumble of data samples by selecting the data subsets those relate to separate models, as well as developed TSK models simultaneously based on the selected data subset. These models may be used for selection of the parameter’s value of a process to achieve a desired goal. Unfortunately, to choose a model that especially suits to a particular test case is a quite difficult task. In order to solve this problem, the user should have to acquire sufficient knowledge of the process through the proposed algorithm.

Table 7

Results of cross-validation among different approaches

Approaches	RMSE value based on training data	RMSE value based on test data
Approach 0	56.92	65.18
Approach I	18.34	26.85
Approach II	7.67	8.11

6. Conclusion

In this paper, a simple and easy to implement algorithm is proposed to detect ambiguities in regression problems using TSK models. The TSK fuzzy models are considered to work in parallel based upon the idea of sharing the data with each other in every step. This method allows the identification of the data that might have been collected from different strategies (or models). We have also demonstrated that the membership functions of the TSK models can be optimized simultaneously using an evolutionary algorithm. The proposed approach is evaluated with artificial data of steering a vehicle to avoid an obstacle using two TSK models and it is applied to a grinding process to select wheel speed in order the achieve the desired surface finish on a work piece.

7. Acknowledgement

The authors are thankful to DAAD (German Academic Exchange Service) for supporting this research work. The authors would also like to thank the anonymous reviewers for their valuable and helpful comments that lead to a significant improvement of the paper.

References:

[1] Yang Z, Gao Y, Zhang D, Huang T (2003) A self-tuning based fuzzy-PID approach for grinding process control. Key Engineering Materials, 238-239: 375-380

[2] Tsang ECC, Yeung DS (1999) Optimizing fuzzy knowledge base by genetic algorithms and neural networks. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Tokyo, Japan

[3] Tönshoff HK, Walter A (1994) Self-tuning fuzzy-controller for process control in internal grinding. J. Fuzzy Sets and Systems, 63(3): 359-373

[4] Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modelling and control. IEEE Trans. Systems, Man, and Cybernetics, 15: 116-132

[5] Sugeno M, Kang GT (1988) Structure identification of fuzzy model. J, Fuzzy Sets and Systems, 28: 15-33

[6] Shen H, Yang J, Wang S (2004) Outlier detecting in fuzzy switching regression models, Lecture Notes in Computer Science, Springer-Verlag GmbH, 3192: 208-215

[7] Simon D (2002) Training fuzzy systems with the extended Kalman filter. J. Fuzzy Sets and Systems, 132(2): 189-199

[8] Satyanarayanan G, Lin IJ, Chen MK (1992) Neural network modelling and multi-objective optimisation of creep feed grinding of super alloys. Int. J. Production Research, 30(10): 2421-2438

[9] Rao KGRK, Kuppuswamy G (1997) Fuzzy logic interface application to the electro-chemical grinding process analysis. Proc. of 17th AIMTDR Conf., REC, Warangal, India, 329-334

[10] Rao AV, Miller D, Rose K, Gersho A (1997) Mixture of experts regressions modeling by deterministic annealing. IEEE Trans. Signal Processing, 45 (11): 2811-2820

[11] Quandt RE, Ramsey JB, (1978) Estimating mixtures of normal distributions and switching regressions, J. American Statistical Association, 73: 730-752

[12] Quandt RE (1972) A new approach to estimating switching regressions, J. American Statistical Association, 67: 306-310

[13] Nandi AK, Pratihar DK (2004) Design of a genetic-fuzzy system to predict surface finish and power requirement in grinding. J. Fuzzy Sets and Systems, 148(3): 487-504

[14] Nandi AK, Banerjee MK (2005) FBF-NN-based modelling of cylindrical plunge grinding process using a GA, J. Material Processing Technology, 162-163C: 655-664

[15] Nagasaka K, Ichihashi H, Leonard R (1995) Neuro-fuzzy GMDH and its application to modelling grinding characteristics. Int. J. Production Research, 33(5): 1229-1240

[16] Ménard M (2001) Fuzzy clustering and switching regression models using ambiguity and distance rejects, Fuzzy Sets and Systems, 122(3): 363-399

[17] Maybeck PS, Pogoda DL (1989) Multiple model adaptive controller for the stol f-15 with sensor/actuator failures, In Proceedings of the 20th Conference on Decision and Control, 1566-1572

[18] Liao TW, Chen JL (1994) A neural network approach for grinding process: modeling and optimization. Int. J. Machine Tools and Manufacture, 34 (7): 919-937

[19] Lee CW, Shin YC (2003) Construction of fuzzy systems using least-squares method and genetic algorithm. J. Fuzzy Sets and Systems, 137(3): 297-323

[20] Klawonn F, Kruse R (1999) Techniques and applications of control systems based on knowledge-based interpolation, In: Leondes C.T. (ed.): Fuzzy theory systems: Techniques and Application, Academic Press, San Diego, 431-460

[21] King RI, Hahn RS (1986) Handbook of modern grinding technology. Chapman and Hall

[22] Kecman V, Pfeiffer B-M (1994) Exploiting the structural equivalence of learning fuzzy systems and radial basis function neural networks, In Proceedings of European Congress on Intelligent Techniques and Soft Computing (EUFIT'94), Achen, 58-66

[23] Herrera F, Lozano M, Verdegay JL (1995) Tuning fuzzy logic controllers by genetic algorithms. Int. J. Approximate Reasoning, 12(3-4): 299-315

[24] Hathaway RJ, Bezdek JC (1993) Switching regression models and fuzzy clustering, IEEE Trans. Fuzzy Systems 1 (3) 195–203

[25] Grauel A, Renners I, Ludwig LA (2000) Optimizing fuzzy classifiers by evolutionary algorithms. In Proceedings of the IEEE 4th International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies

[26] Goel P, Dedeoglu G, Roumeliotis SI, Sukhatme GS (2000) Fault Detection and Identification in a Mobile Robot using Multiple Model Estimation and Neural Network, IEEE International Conference on Robotics and Automation (ICRA), San Fransisco, CA, Apr 2000

[27] Figueiredo M, Gomide F (1999) Design of fuzzy systems using neurofuzzy networks. IEEE Trans. Neural Networks, 10: 815–827.

[28] Cheng HD, Chen JR (1997) Automatically determine the membership function based on the maximum entropy principle. Information Sciences, 96(3-4): 163-182

[29] Bai A (2003) Determining fuzzy membership functions with Tabu search—an application to control. J. Fuzzy Sets and Systems, 139(1): 209-225

[30] Arslan A, Mehmet K (2001) Determination of fuzzy logic membership functions using genetic algorithms. J. Fuzzy Sets and Systems, 118(2): 297-306

[31] Abonyi J, Babuska R, and T. Chovan T, Szeifert F (2000) Incorporating prior knowledge in fuzzy c-regression models - Application to system identification, Proceedings of the Intelligent Systems in Control and Measurement Symposium (INTCOM 2000), 99-110