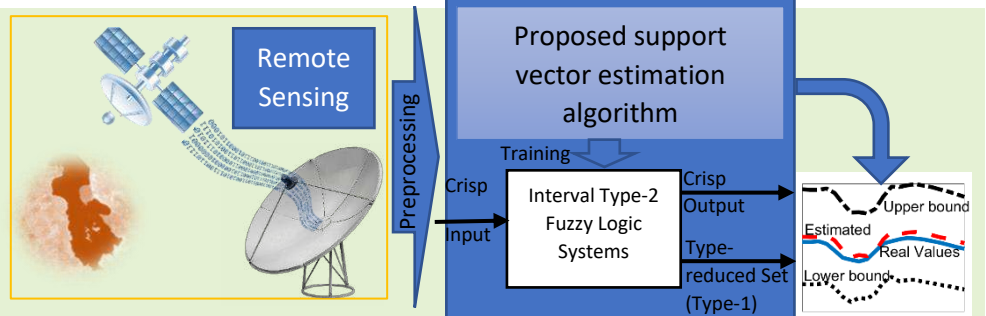


Prediction Interval Identification Using Interval Type-2 Fuzzy Logic Systems: Lake Water Level Prediction Using Remote Sensing Data

M. A. Khanesar¹, Senior Member, IEEE, David T. Branson²

Abstract—This paper presents a novel approach to identify the prediction interval associated with data using interval type-2 fuzzy logic systems with support vector regression. For such a purpose, a constrained quadratic objective function is defined which is then solved using well-established quadratic programming approaches. Not only does the output of interval



type-2 fuzzy logic system replicates the measured value, but also it provides the lower bound and the upper bound for measured data values. In the proposed approach, to have more valuable information, a penalty term is added in the cost functions to have full control over the width of prediction interval. This method has been successfully applied to two benchmark identification problems. It is observed that by using the control parameter in the cost function, it is possible to obtain a narrower, yet inclusive prediction interval. Furthermore, superior prediction accuracy is obtained compared to existing methods in literature. Motivated by these results, the proposed approach is used to predict time series collected using a satellite from Urmia lake water level which resulted in high accuracy and an inclusive prediction interval. The graphical abstract presented for the paper illustrates the overall data gathering as well as data analysis made to estimate the prediction interval associated with Urmia lake water level data.

Index Terms—Fuzzy neural networks, prediction algorithms, Estimation, prediction interval, remote sensing, level measurement,

I. INTRODUCTION

Interval type-2 fuzzy logic systems (IT2FLSs) are widely known to be an inevitable option in the presence of high levels of uncertainties and noise in the system [1]. In a fuzzy logic system, the membership grade assigned to an input by various experts may be different resulting in interval membership grades. Interval type-2 fuzzy membership functions (MFs) are a promising method to deal with such different expert knowledge which benefit from an infinite number of type-1 fuzzy MFs. Although use of interval type-2 fuzzy MFs makes the structure of IT2FLSs more complicated, it allows them to deal with high levels of uncertainties in the system. The uncertainty can exist in the MFs and/or in the consequent part resulting in a histogram of values in the consequent part [2].

“This work is funded and supported by the Engineering and Physical Sciences Research Council (EPSRC) under grant number: EP/R021031/1 - New Industrial Systems: Chatty Factories.”

¹ M. A. Khanesar, is with the Faculty of Engineering, University of Nottingham, Nottingham, NG7 2RD, U.K. (e-mail: eezma5@exmail.nottingham.ac.uk).

² David T. Branson is with the Faculty of Engineering, University of Nottingham, Nottingham, NG7 2RD, U.K. (e-mail: David.Branson@nottingham.ac.uk).

The additional degree of freedom which exists in the structure of IT2FLSs make it possible to deal with uncertainty and noise which may exist in real world data.

Real world datasets inherently suffer from nonlinearity, uncertainty and noise. Although fuzzy logic systems, especially in the case of Mamdani type, are known to be general function approximators, minimum functional approximation error inevitably exist [3]. A bounding interval for data can be more useful than just identifying the crisp output value of the system as it guarantees the prediction interval for future data samples. Not only can IT2FLSs deal with uncertainty and noise but also they promise a prediction interval that covers data. Piecewise linear methods as well as interval fuzzy models have already been used to find prediction intervals [4,5]. A linear programming approach is used to estimate the parameters of these structures [4,5]. However, the methods investigated in these studies do not provide any means to control the prediction interval width, which is the primary motivation of this study.

Finding an unviolated prediction interval covering all future possible values is a very useful practice. General function approximators such as fuzzy logic models may be used to

estimate interval values associated with data. To show the importance and place of prediction interval identification, several possible engineering applications that already exist in literature are studied. For instance, when modeling a nonlinear circuit, the elements of the circuit that suffer from uncertainties in their real values due to manufacturing variations. Furthermore, environmental conditions may impose more uncertainties to their characteristic [5]. Characteristic identification of physical systems may be useful for fault identification as well. In [6], the physical system is modeled completely in terms of an IT2FLS, the violation of the prediction interval associated with system is an indication of fault in the system. It is further possible to use multiple IT2FLSs describing various faulty conditions of the system which can be used to identify the cause of the fault rather than just its occurrence [7]. To achieve stable operation of an energy management system, it is required to utilize knowledge about the predicted values as well as uncertainty associated with wind speed, solar energy and other resources used in such an energy management system. Such resources are highly stochastic. A robust energy management system using wind power and speed interval prediction is investigated in [8], where the obtained prediction interval for wind is used for management purposes. Another application of fuzzy prediction interval approach is in stock exchange price prediction [9]. For such a purpose, simulation results show that an IT2FLS whose parameters are trained using genetic algorithms can successfully identify the prediction interval associated with data [9].

Various algorithms have been used to estimate the parameters of IT2FLSs. Such optimization algorithms can be placed into three main categories: derivative-based optimization methods, derivative-free optimization approaches and hybrid algorithms [10]. From another point of view, training methods for IT2FLSs can be placed into two categories: iterative approaches and non-iterative approaches [11]. While optimizing IT2FLSs using least square is a non-iterative approach, gradient based methods instead require several iterations before converging to their optimal solutions [12]. Support vector regression (SVR) is an alternative non-iterative training machine learning method that can successfully be applied to train IT2FLSs. It is widely known that an IT2FLS trained using this method results in superior generalization properties than most of other approaches [13]. In [13,14], SVR is used to estimate the parameters of an IT2FLS, the comparison results show the superiority of this system over several existing system identification methods. A complete review summarizing various parameter estimation methods for IT2FLSs is provided in [10]. The proposed approach in this study is a non-iterative derivative-free method with a few design parameters which makes it an ideal choice when dealing with nonlinear interval system identification. Being non-iterative, the probability of entrapment in a local minimum for the proposed algorithm is minimum as well.

Statistical approaches such as bootstrap, Bayesian method as well as Kalman filter have previously been applied to find the confidence interval associated with data [15,16]. The structures used in these studies are mostly artificial neural networks (ANNs) and cover a wide range of applications including

decision making [8] and prediction interval estimation methods [17]. However, these approaches mostly involve the calculation of the Hessian value of an ANN structure's output with respect to its parameters which is known to be time consuming. Moreover, the identification error in these cases needs to be zero mean with a normal probability distribution function [18].

The fuzzy logic system parameter estimation methods used in [4,5,19] rely on a constrained cost function. A constrained least square method is formulated to find the fuzzy prediction interval whilst having measured data between bounds defined by fuzzy model [19]. However, such a prediction interval estimator did not have any parameter to control the width of the prediction interval. This is the main drawback of the algorithm introduced in [19] which may result in too conservative prediction interval. The main motivation of this paper is to have full control over the width of prediction interval to obtain a narrow, yet inclusive one.

In this study, the original version of SVR, a powerful machine learning approach, for the training of IT2FLSs [13,14] is modified. The two main motivations for the modification made in this study are:

- 1) To deal with identification problems when the estimation error is not Gaussian.
- 2) To have control over width of the prediction interval

This modified version of SVR includes terms corresponding to the width of the prediction interval associated with system output in the cost function to control it. These modifications form the main contribution of the proposed approach over existing approaches. Although the prediction interval covering data presents valuable information about it, too wide prediction interval contains less specific information about data. It is therefore highly desirable to have some means to control the width of prediction interval, which in turn adds another objective function, making the parameter estimation of the IT2FLS a multi-objective optimization problem. The added objective function includes the width of the output interval associated with fuzzy system output. Two more constraints are required to be added to make sure that measured data samples do not fall outside the interval defined by the IT2FLS. The fact that the proposed approach includes some means to control width of prediction interval found by the fuzzy system is the main contribution of the current paper over previously studied SVR approach in [13,14]. Similar to previous approaches such as [13,14], the optimization problem is solved using a quadratic programming algorithm. Uncertainty analysis in the case of statistical approaches necessitates the calculation of a Hessian matrix which is a very complex task. Furthermore, these approaches suffer from requirements on error such as the normal distribution function and zero-mean which may not be valid in some cases. The proposed approach is an algebraic approach which relaxes some of the assumptions about error such as being zero mean and having a normal probability distribution function. The proposed algorithm is then used to estimate the prediction interval for two interesting benchmark problems. Simulation results support the fact that using the proposed approach, one can have full control over the width of the prediction interval, while maintaining prediction accuracy.

Earth observation has different branches such as ground-sensing networks and satellite remote sensing [20]. The data collected using satellites can be used to gain a wide range of knowledge about the social, economic, and environmental condition of the region. Spatial temporal dynamic analysis is possible through remote sensing. For instance, urban population in an area is estimated by analyzing the nighttime light. The brightness of an area in this case is an indication of its population, wealth and similar parameters associated with the region [21]. Haze pollution monitoring is another application of satellite remote sensing [22] which results in high quality measurements. Various machine learning techniques including texture analysis [23], clustering and classification [24], as well as visualization, analysis and interpretation [25] are among the approaches which have already been applied to the remote sensing dataset. The analysis of water resources is crucial because of its social effect on the life of people living in the nearby area. Such analysis is possible through remote sensing using the data collected from satellites¹. Urmia lake has been paid much attention in previous studies due to its high social impact on life of people neighboring area and its environmental conditions [26,27]. In this study, the proposed multi-objective prediction interval is used to analyze the time series associated with the water level of Urmia lake. Comparisons are provided between the proposed approach and state-of-the-art studies applied to the same dataset which show the superior performance of the proposed approach.

This paper is organized as follows: In Section II, a literature survey is performed addressing relevant approaches. An overview of the basic structure of the IT2FLS is provided in Section III. The proposed methodology for the training of IT2FLSs are presented in Section IV. The experimental results of the proposed IT2FLSs are illustrated in Section V. Finally, in Section VI, the concluding marks are presented.

II. LITERATURE SURVEY

In this section, a list of articles related to the estimation of the prediction interval associated with data are presented. Existing approaches can be placed into three major categories of classical statistical approaches, upper and lower estimation methods and other optimization-based approaches which does not include extraction of upper and lower bound of data *a priori* (see Fig. 1).

A. Classical Statistical Approaches

The first class are classical statistical approaches that rely on uncertainty analysis to find the standard deviation associated with the output of model. These methods assume a normal distribution function for error which makes it possible to come up with an appropriate confidence interval covering data.

The bootstrap method aims at constructing several subsets of data by re-sampling the original dataset. Each subset is then modeled independently using an identifier, and outputs of trained models combined to forecast data. The most straightforward method is to use the average value of constructed

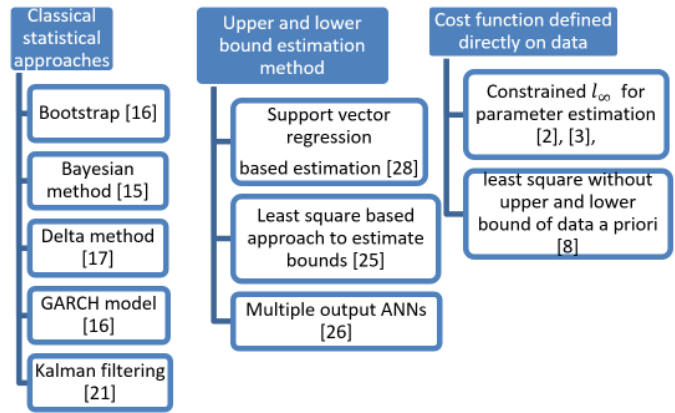


Fig. 1. Categories of different estimation algorithms for prediction interval identification

models as the overall model. However, Bayesian model averaging as well as weighted average least square can be named as alternative approaches [28]. Bayesian model averaging assigns different gains to each model such that the best model has the greatest share in the output and the worst one has the least impact on it. Overall model averaging in this case results in a model that can outperform every single input model. On the other hand, the weighted average least square method uses orthogonalization which reduces the computational burden whilst improving the accuracy of the system [28].

Among various model averaging methods, the simplest one is taking the average model of several ANNs previously investigated in [29]. This approach is examined on the energy market of Victoria region in Australia and New York city with a sampling time of 30 minutes [29]. In this work, time-varying variance associated with models is estimated using the GARCH model. This variance can be controlled to reflect the prediction interval associated with the $(1-\alpha)\%$ confidence interval. The fundamental assumption for the maximum likelihood estimator required for training a GARCH model is the normal distribution of error.

Bayesian approach is another statistical method that can be used to find the prediction interval associated with data. The parameters in this case are estimated such that they minimize a regularized cost function that includes the sum of squared error as well as the norm of network weights. The uncertainty analysis in this algorithm involves the estimation of a Hessian matrix that can be cumbersome in terms of ANNs as this structure has a large number of parameters [18].

The Delta method, another statistical approach to estimate the prediction interval, uses a Taylor expansion of a neural network to find the uncertainty associated with its output. This algorithm suffers from computational burden due to dependency of its uncertainty analysis on a Hessian matrix [30].

Boot-strap, Bayesian approach and delta methods to find prediction interval using ANNs are comprehensively investigated in [18] for ten different case studies. The investigated datasets cover a wide range including synthetic dataset, body fat estimation, real-world baggage handling system, concrete compressive strength and six more datasets. It is concluded

¹Lake Urmia (0115) Height Variations from TOPEX POSEIDON and Jason series Altimetry

that Bayesian method is the most reliable prediction interval in most cases. However, the computational burden associated with this method may be against wide-spread use of this algorithm [18].

In the case of using the least square estimation method for ANN parameters, the covariance matrix associated with this estimator can be constructed based on its properties [31]. In [31], numerical analysis is presented to demonstrate the quality of the prediction interval of ANNs for an artificial plant as a benchmark example. However, the least square algorithm can be applied on a limited class of linear regression models as well as the output layer of ANNs when a linear activation function is used in the output layer. The generalization of this approach to a wider class of intelligent structures would be difficult and may involve approximation.

The Kalman filter family are well-known estimation methods that can be effectively used to tune the parameters of artificial intelligent structures [32]. Its extended version benefits from Taylor expansion and can be used to estimate the parameters of ANNs when they appear non-linearly in ANN output. There exist various versions of Kalman filtering including the extended Kalman filter with U-D factorization, designed to improve the numerical stability of the estimation method [15]. The adaptive Kalman filter benefits from mechanisms to tune the Kalman filter parameters during training automatically. This will contribute to having less design parameters and consequently less design iterations. On the other hand, their decoupled extended Kalman filters are frequently used due to its decoupling procedure. However, they impose more approximation to reduce the size of covariance matrix [33].

Uncertainty in data directly influences the innovation covariance matrix. Standard deviation associated with each output can be calculated from the diagonal elements of this matrix resulting in finding prediction intervals associated with each output. Different members of the Kalman filter family has been previously applied to estimate the prediction interval associated with market clearing price [15] as well as short-term load forecast [34].

The main disadvantage of using the Kalman filter family to estimate the confidence intervals is that the assumptions on noise including the Gaussian probability distribution function and zero-mean are among the assumptions required for this estimator. Moreover, the Taylor expansion needed to estimate the parameters that appear nonlinearly in the output imposes more inaccuracy to the approximate method as its higher order terms must be neglected [33].

Fuzzy systems may be considered as alternative general function approximators for ANNs, and may outperform ANNs in some cases [35], [36]. The fuzzy confidence interval is developed in [37] using statistical methods under general statistical assumptions for data such as zero-mean error, and the normal probability distribution function for error.

B. Lower Upper Bound Estimation Algorithm

Ideal absolute and relative lower and upper bounding of data can be generated to identify the prediction interval associated with it [38]. Multiple linear regression models benefiting from

the least square algorithm to estimate their parameters are used in [38] to identify the prediction interval width as well as the measured value. This approach is used to find the prediction interval of the daily-sampled discharge value of the Yangtze river, the longest river in Asia, located within the Chinese territory [38]. The lower upper bound estimator (LUBE) proposed by Khosravi et. al [39] uses an ANN with two outputs that directly replicate the upper and lower bounds. A cost function that includes the coverage probability as well as the width of the prediction interval is used to estimate the parameters of the ANN with two outputs [39]. The main disadvantage of this method is the requirement to select design parameters in the cost function that greatly influence its performance. To successfully implement this algorithm, several iterations with different parameter values in the cost function may be required before ending up with an appropriate selection. A similar two output ANN is used in [40] that benefits from fuzzy objective functions to reduce the number of design parameters in its cost function. This fuzzy objective function controls the training procedure of the ANN with two outputs. Generally, the LUBE methods alleviate the requirement for the error probability distribution to be normal which is common for statistical approaches and limits their applicability [39] and [40].

In a study performed by N. Shrivasta *et. al.*, electricity price was identified using radial basis neural networks [41]. In this work two support vector machines are used to train the radial basis kernels to follow the trend of lower and upper bounds of data. The parameters associated with the two support vector machines are selected using particle swarm optimization rather than grid partitioning to speed up the overall estimation method. The main objective function considered for the particle swarm optimizer includes a term associated with coverage and width criterion. The studied method successfully obtained the prediction interval for electricity prices of the Ontario electricity market an hour-ahead forecasting as well as three-hours ahead forecasting basis. The prediction interval associated with the Pennsylvania–New Jersey– Maryland day-ahead market, and real-time market is considered in this study in a daily basis. It is important to demonstrate that the proposed approach is fast enough to perform prediction in a timely manner. Simulation results demonstrate that the prediction algorithm is reasonably fast, as it takes a couple of minutes to be completed, that makes it suitable for the prediction of a few hours ahead [41].

The simplicity of the LUBE method is its main advantage and has resulted in the use of this algorithm in several problems including wind power forecasting [40], flood prediction interval forecast [42], wind speed prediction interval [43] tidal current forecast [44], and short term photo voltaic forecast [45].

The main disadvantage is the need to find the upper and lower bound *a priori* as well as the selection of parameters in the cost function. Although the fuzzy approach in [40] decreases the number of design parameters in the cost function, finding appropriate upper and lower bounds is still required for successful use. The approach presented in this study, does not require the knowledge of the upper and lower bound associated

with data *a priori*. This is the main improvement offered by the proposed algorithm over approaches studied under the LUBE title.

C. Methods using Cost Function Defined Directly on Data

In the approach investigated in [4,5], two lower and upper fuzzy models are considered to identify the prediction interval. Data needs to be between the upper and the lower fuzzy models while the outputs of the fuzzy models are as close as possible to the measured data. Hence, the problem is formulated as a constrained optimization problem. The cost function considered in this study is the l_∞ -norm of the difference between the output of the two fuzzy models and the measured data under the constraint imposed by the requirement to have data in the interval defined by the lower and upper fuzzy models. Similar methods are investigated for the identification of uncertain systems for robust fuzzy control purposes, where the lower and upper fuzzy models are used to identify uncertain dynamics of nonlinear systems in state space form. The identified interval fuzzy model is then used to design a robust controller for an inverted pendulum and a bulk converted circuit [46]. In [4,5,46] a linear programming approach is used to find the upper and lower fuzzy models. Other than application to robust control, the fuzzy prediction interval has been successfully applied to fault detection problems [47].

Other than the interval associated with the output of a fuzzy model, in [19] another approach is proposed which estimates the interval associated with the consequent part parameters as well using the least square method. Hence, this approach is not a black box method as it gives more information about the internal parameters of the system.

The main advantage of the approaches in this category over statistical approaches is that they do not depend on the assumption on estimation error to have normal probability distribution function. They also do not necessitate finding the lower and upper bound for data *a priori*, a common requirement of the LUBE method. However, the downfall of the algorithms used in [4,5,19,46] is that no control exists over the width of the prediction interval which may result in an unnecessarily large width for the prediction interval meaning it contains less specific information about data. Motivated by this shortcoming an algorithm is proposed in this study that falls within algorithms explained in Section II-C and is capable of controlling the width of prediction interval.

D. General overview of aforementioned methods

The general overview of the proposed approaches is presented here. Classical statistical approaches are widely used for uncertainty analysis associated with the prediction interval. They benefit from statistical analysis which makes them a reliable approach. However, the assumptions associated with these approaches including the Gaussian probability distribution function for the prediction error as well as being zero-mean are among the restricting requirements for this method. The LUBE is the second approach investigated in this paper. The interesting feature of this algorithm is that

after finding an absolute or relative upper and lower bound for data, finding the nonlinear function which can approximate them is straightforward. However, considering the fact that uncertainty may not be distributed uniformly makes finding the lower and upper bound associated with data *a priori* very difficult. The main disadvantage of the methods investigated in Section II-C is that no control exists over the width of the prediction interval in these algorithms which may result in a wider prediction interval than required. If the width of prediction interval is wider than required it contains less specific information about data which needs to be avoided. The aforementioned shortcomings associated with the three categories of algorithms are the main motivation for the proposed approach in this paper.

III. GENERAL STRUCTURE OF INTERVAL TYPE-2 FUZZY LOGIC SYSTEM

The general structure of interval type-2 fuzzy systems including its main building blocks is depicted in Fig. 2. Several structures for IT2FLSs and its type-reducers are investigated in [48,49]. The structure used in this study benefits from interval type-2 fuzzy MFs in the antecedent part and interval values for the consequent part parameters. A typical fuzzy IF-THEN rule for such a structure is as follows.

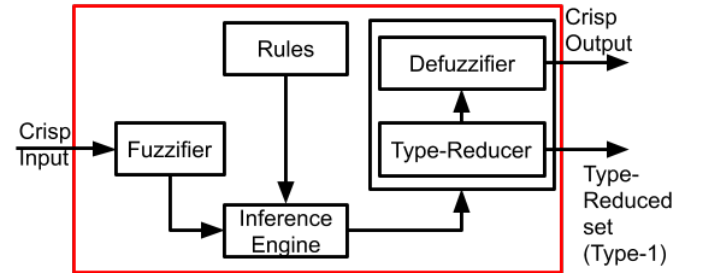


Fig. 2. Type-2 fuzzy logic system building blocks

IF x_1 is \tilde{A}_{j1} and x_2 is \tilde{A}_{j2} and... and x_n is \tilde{A}_{jn}

$$THEN \quad y_j = \sum_{i=1}^n \tilde{\alpha}_{ij} x_i + \tilde{\beta}_j \quad (1)$$

where x_1, x_2, \dots, x_n are the input variables, y is the single output variable. Moreover, \tilde{A}_{ij} 's are interval type-2 fuzzy MFs for the j^{th} rule of the i^{th} input. $\tilde{\alpha}_{ij}$ and $\tilde{\beta}_j$ ($i = 1, \dots, n, j = 1, \dots, M$) are the interval parameters in the consequent part of the rules that satisfy the following equation.

$$\tilde{\alpha}_{ij} \in [\underline{\alpha}_{ij}, \bar{\alpha}_{ij}], \quad \tilde{\beta}_j \in [\underline{\beta}_j, \bar{\beta}_j] \quad (2)$$

The following definitions are made.

$$\underline{F}_j = \sum_{i=1}^n \underline{\alpha}_{ij} x_i + \underline{\beta}_j \quad (3)$$

$$\bar{F}_j = \sum_{i=1}^n \bar{\alpha}_{ij} x_i + \bar{\beta}_j \quad (4)$$

$$\underline{w}^j(x) = \underline{\mu}_{\bar{F}_1}(x_1) * \dots * \underline{\mu}_{\bar{F}_n}(x_n) \quad (5)$$

$$\bar{w}^j(x) = \bar{\mu}_{\bar{F}_1^j}(x_1) * \dots * \bar{\mu}_{\bar{F}_n^j}(x_n) \quad (6)$$

where $\underline{\mu}_{\bar{F}_k^j}(x_k)$ are $\bar{\mu}_{\bar{F}_k^j}(x_k)$ are the lower and upper MF corresponding to j^{th} rule for x_k and "*" is a t-norm operator. The output value of IT2FLS is given as

$$Y(x) = [y_l(x), y_r(x)] \quad (7)$$

where $x \in R^n$ is a vector of inputs of system. Among various defuzzifications approaches to calculate the output of an IT2FLS, the Maclaren based first order approximate one is chosen [49]. The accuracy of this defuzzification algorithm is lower than the exact model of the enhanced Karnik-Mendel model approach [50] and higher than the approximate models of Biglarbegian-Melek-Mendel [51] and Nie-Tan [52]. The computational burden for such an algorithm is less than the enhanced Karnik-Mendel model as it does not necessitate the sorting procedure required by it. The Maclaren series expansion based first order approximate output of the IT2FLS is as follows [49]:

$$y \in [y_l, y_r] \quad (8)$$

where y_l and y_r are the left most and right most values of output of IT2FLSs, respectively. These parameters are calculated as follows.

$$y_r \approx \frac{\sum_{j=1}^M (\bar{w}^j + \underline{w}^j) \bar{F}^j + \sum_{j=1}^M (\text{sign}(\bar{m}^j) \Delta w^j \bar{F}^j)}{\sum_{j=1}^M (\bar{w}^j + \underline{w}^j) + \sum_{j=1}^M (\text{sign}(\bar{m}^j) \Delta w^j)} \quad (9)$$

where:

$$\bar{m}^j = \bar{F}^j - \frac{\sum_{j=1}^M \bar{w}^j \bar{F}^j}{\sum_{j=1}^M \bar{w}^j} \quad (10)$$

and $\Delta w^j = \bar{w}^j - \underline{w}^j$. Furthermore, y_l is calculated as

$$y_l \approx \frac{\sum_{j=1}^M (\bar{w}^j + \underline{w}^j) \underline{F}^j - \sum_{j=1}^M (\text{sign}(\underline{m}^j) \Delta w^j \underline{F}^j)}{\sum_{j=1}^M (\bar{w}^j + \underline{w}^j) - \sum_{j=1}^M (\text{sign}(\underline{m}^j) \Delta w^j)} \quad (11)$$

where:

$$\underline{m}^j = \underline{F}^j - \frac{\sum_{j=1}^M \underline{w}^j \underline{F}^j}{\sum_{j=1}^M \underline{w}^j} \quad (12)$$

The final crisp output value of IT2FLS is obtained as

$$Y(x) = \frac{y_l + y_r}{2} \quad (13)$$

It is then possible to rewrite (9) as

$$y_r = \sum_{j=1}^M \nu_R^j \bar{F}_R^j \quad (14)$$

where:

$$\nu_R^j = \frac{\bar{w}^j + \underline{w}^j + \text{sign}(\bar{m}^j) \Delta w^j}{\sum_{j=1}^M (\bar{w}^j + \underline{w}^j) + \sum_{j=1}^M (\text{sign}(\bar{m}^j) \Delta w^j)} \quad (15)$$

The parameter y_r in a vector form is obtained as:

$$y_r = \phi_R \bar{\theta} \quad (16)$$

where

$$\phi_R = [\bar{\nu}_R^T, \bar{\nu}_R^T x_1, \dots, \bar{\nu}_R^T x_n]^T \quad (17)$$

and $\bar{\alpha}_R$ is defined as

$$\bar{\nu}_R = [\nu_R^1, \dots, \nu_R^M]^T \quad (18)$$

Furthermore, $\bar{\theta}$ is defined as

$$\bar{\theta}_{(n+1).M} = [\bar{\beta}_1, \dots, \bar{\beta}_M, \bar{\alpha}_{11}, \dots, \bar{\alpha}_{1M}, \dots, \bar{\alpha}_{n1}, \dots, \bar{\alpha}_{nM}]$$

where $\bar{\alpha}$'s and $\bar{\beta}$'s are the consequent part parameters defined in (4). Similarly, it is possible to rewrite the equation corresponding to y_l (11) in a vector form as

$$y_l = \sum_{j=1}^M \nu_l^j \underline{F}_l^j \quad (19)$$

where:

$$\nu_l^j = \frac{(\bar{w}^j + \underline{w}^j) - (\text{sign}(\underline{m}^j) \Delta w^j)}{\sum_{j=1}^M (\bar{w}^j + \underline{w}^j) - \sum_{j=1}^M (\text{sign}(\underline{m}^j) \Delta w^j)} \quad (20)$$

$$y_l = \phi_L \underline{\theta} \quad (21)$$

where:

$$\phi_L = [\bar{\nu}_L^T, \bar{\nu}_L^T x_1, \dots, \bar{\nu}_L^T x_n]^T \quad (22)$$

and $\bar{\alpha}_L$ is defined as.

$$\bar{\nu}_L = [\nu_L^1, \dots, \nu_L^M]^T \quad (23)$$

Furthermore, $\underline{\theta}$ is defined as.

$$\underline{\theta}_{(n+1).M} = [\underline{\beta}_1, \dots, \underline{\beta}_M, \underline{\alpha}_{11}, \dots, \underline{\alpha}_{1M}, \dots, \underline{\alpha}_{n1}, \dots, \underline{\alpha}_{nM}]$$

where $\underline{\alpha}$'s and $\underline{\beta}$'s are the consequent part parameters defined in (3).

IV. SUPPORT VECTOR REGRESSION METHOD

Support vector regression is a powerful machine learning approach that is widely used to estimate the parameters of fuzzy logic systems including IT2FLSs [13,14]. Let N be the number of training data samples to train IT2FLSs $\{(x_1, t_1), \dots, (x_N, t_N)\}$, with $t_k, k=1, \dots, N$ being the target values for IT2FLSs. The SVR method is designed to guarantee that training error never exceeds ε [13,14]. Such behavior is similar to a dead zone term, $D\varepsilon_k$, which can add a penalty term to the cost function when an error is larger than ε and stays neutral otherwise. A dead-zone function is as follows:

$$D\varepsilon_k = \begin{cases} 0 & \text{if } |e_k| \leq \varepsilon \\ |e_k| - \varepsilon & \text{otherwise} \end{cases}, \quad k=1, \dots, N$$

where e_k is the identification error. The constrained optimization problem to estimate the parameters of the IT2FLS is as follows [13,14]:

$$\min_{\underline{\theta}, \bar{\theta}, \xi, \xi^*} \frac{1}{2} \bar{\theta}^T \underline{\theta} + \frac{1}{2} \bar{\theta}^T \bar{\theta} + C \sum_{k=1}^N (\xi_k + \xi_k^*) \quad (24)$$

$$\text{s.t.} \quad t_k - \frac{1}{2} (\phi_{L,k} \underline{\theta} + \phi_{R,k} \bar{\theta}) \leq \varepsilon + \xi_k, \quad k=1, \dots, N \quad (25)$$

$$\frac{1}{2} (\phi_{L,k} \underline{\theta} + \phi_{R,k} \bar{\theta}) - t_k \leq \varepsilon + \xi_k^*, \quad k=1, \dots, N \quad (26)$$

$$\xi_k, \xi_k^* \geq 0 \quad \forall k \quad (27)$$

In the case when $\varepsilon < |e_k|$, the terms ξ_k and ξ_k^* act as a penalty term for cost function. The parameter C forms a trade-off between the model complexity and its accuracy.

This cost function and its solution using linear programming approaches have been previously considered in literature [13,14]. Although such approaches result in high performance identifiers with high generalization performances, there is no control over the width of the interval value of the output that is equal to $(y_r - y_l)$ point-wise. As mentioned earlier, it is highly desired to have a narrow interval width as it contains more information about data. Motivated by this fact, a modified cost function for the SVR model is proposed in this study in Section 5.

V. PROPOSED MULTI-OBJECTIVE SUPPORT VECTOR REGRESSION METHOD

As mentioned earlier, the algorithm proposed in this study not only identifies crisp output value, but also the prediction interval associated with data. The proposed method benefits from penalty terms which provide the means to control the width of the prediction interval using an appropriate design parameter. The mathematical formulation of the constrained optimization functions as well as the corresponding quadratic programming approaches to solve these problems are presented in this Section.

A. Cost Function Formulation

The objective is to find the nonlinear interval function covering the data plus having the type reduced output of the IT2FLS approximate the measured data points. It is desirable to allow control of the output interval width which is equal to $(y_r - y_l)$ point-wise. Therefore, the constrained optimization problem is modified as follows:

$$\min_{\underline{\theta}, \bar{\theta}, \xi, \xi^*} \frac{1}{2} \underline{\theta}^T \underline{\theta} + \frac{1}{2} \bar{\theta}^T \bar{\theta} + C \sum_{k=1}^N (\xi_k + \xi_k^* + \xi_k^- + \xi_k^+) + \gamma \underbrace{\sum_{k=1}^N (t_k - \phi_{L,k} \underline{\theta})^2}_{I_1} + \gamma \underbrace{\sum_{k=1}^N (t_k - \phi_{R,k} \bar{\theta})^2}_{I_2} \quad (28)$$

$$s.t. \quad t_k - \frac{1}{2}(\phi_{L,k} \underline{\theta} + \phi_{R,k} \bar{\theta}) \leq \varepsilon + \xi_k, \quad k = 1, \dots, N, \quad (29)$$

$$\frac{1}{2}(\phi_{L,k} \underline{\theta} + \phi_{R,k} \bar{\theta}) - t_k \leq \varepsilon + \xi_k^*, \quad k = 1, \dots, N, \quad (30)$$

$$\phi_{R,k} \bar{\theta} - t_k \leq \varepsilon + \xi_k^+, \quad k = 1, \dots, N, \quad (31)$$

$$t_k - \phi_{L,k} \underline{\theta} \leq \varepsilon + \xi_k^-, \quad k = 1, \dots, N, \quad (32)$$

$$\xi_k, \xi_k^*, \xi_k^+, \xi_k^- \geq 0 \quad \forall k \quad (33)$$

The newly added terms and constraints with respect to the existing SVR method [13,14] are the terms I_1 , and I_2 terms in (28) as well as the constraints of (32-34). The nonequality of (31), if fulfilled, guarantees that the target values do not exceed the y_r . On the other hand, the nonequality of (32), if fulfilled, guarantees that the target values do not fall below y_l . The terms I_1 , I_2 are used to control the width of interval

with γ being a tuning parameter. While a large value for γ may decrease the prediction accuracy, a small value for it may result in a wide interval value for the output that includes less valuable information.

A Maclauren series first order approximation of type-reduction + defuzzification, previously designed in [49], is used. The regressor values $\phi_{L,k}$ and $\phi_{R,k}$ depend on the consequent part parameter values that in turn result in a two step optimization. In the first step, since the consequent part parameters are unknown, the regressor values ϕ_L and ϕ_R are chosen as:

$$\phi_R = [\vec{\nu}_R^T, \vec{\nu}_R^T x_1, \dots, \vec{\nu}_R^T x_n]^T \quad (34)$$

and $\vec{\nu}_R$ is defined as:

$$\vec{\nu}_R = [\nu_R^1, \dots, \nu_R^M]^T \quad (35)$$

where:

$$\nu_R^j = \frac{\bar{w}^j}{\sum_{j=1}^M \bar{w}^j} \quad (36)$$

and

$$\phi_L = [\vec{\nu}_L^T, \vec{\nu}_L^T x_1, \dots, \vec{\nu}_L^T x_n]^T \quad (37)$$

and $\vec{\nu}_L$ is defined as:

$$\vec{\nu}_L = [\nu_L^1, \dots, \nu_L^M]^T \quad (38)$$

where:

$$\nu_L^j = \frac{\underline{w}^j}{\sum_{j=1}^M \underline{w}^j} \quad (39)$$

In the second step, based on the estimations made in the first step for $\underline{\theta}$ and $\bar{\theta}$, the newer values of regressors for the IT2FLS, ϕ_L and ϕ_R , are calculated considering (17), (18), (22) and (23). The pseudocode of the proposed algorithm is as follows:

- 1) Input Selection and data processing
- 2) Present data in terms of the maximum and minimum
- 3) Split the data to test and train dataset
- 4) MF Generation for the IT2FLS
- 5) Obtain the regressors using (34) - (39).
- 6) Tune the consequent part parameters (first stage) to obtain the regressors using the quadratic programming approach to solve optimization problem (28)-(33)
- 7) Obtain the regressors using the updated consequent part parameters using (15), (17), (18), (20), (22) and (23).
- 8) Tune the consequent part parameters (second stage) using the quadratic programming approach to solve (28)-(33)
- 9) Evaluate the performance for the train and test data. If error is satisfactory STOP, otherwise GOTO 4).

The formulation of this problem in terms of quadratic programming is given in the following section.

B. Solution to Cost Function in terms of a Quadratic Programming Approach

A quadratic programming problem is defined as follows and can be solved using various commercially available software such as the Matlab[®] quadprog command.

$$\min_x \frac{1}{2} x^T H x + f^T x \quad (40)$$

provided that:

$$Ax \leq b \quad (41)$$

where H and A are matrices and f , b , and x are vectors.

Formulation of the problem mentioned in Section V.A in terms of quadratic programming requires the definition of an unknown parameter x as the vector of unknowns as follows.

$$x = [\underline{\theta} \quad \bar{\theta} \quad \xi_k \quad \xi_k^* \quad \xi_k^- \quad \xi_k^+] \quad (42)$$

The matrix H is presented as follows.

$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \quad (43)$$

with its components being represented as follows.

$$H_{11} = \begin{bmatrix} I + \gamma \phi_{L,1} \phi_{L,1}^T \\ \vdots \\ I + \gamma \phi_{L,k} \phi_{L,k}^T \\ \vdots \\ I + \gamma \phi_{L,n+1} \phi_{L,n+1}^T \end{bmatrix}^T$$

$$H_{11} \in R^{(n+1) \cdot M \times (n+1) \cdot M}$$

$$H_{22} = \begin{bmatrix} I + \gamma \phi_{R,1} \phi_{R,1}^T \\ \vdots \\ I + \gamma \phi_{R,k} \phi_{R,k}^T \\ \vdots \\ I + \gamma \phi_{R,n+1} \phi_{R,n+1}^T \end{bmatrix}^T$$

$$H_{22} \in R^{(n+1) \cdot M \times (n+1) \cdot M}$$

$$H_{12} = H_{21} = O_{(n+1) \cdot M \times (n+1) \cdot M}$$

$$H_{13} = O_{(n+1) \cdot M \times 4N}$$

$$H_{23} = O_{(n+1) \cdot M \times 4N}, \quad H_{31} = O_{4N \times (n+1) \cdot M}$$

$$H_{32} = O_{4N \times (n+1) \cdot M}, \quad H_{31} = O_{4N \times 4N}$$

The vector f is represented as follows.

$$f = [f_1 \quad f_2 \quad f_3] \quad (44)$$

where f_1 , f_2 and f_3 are represented as follows.

$$f_1 = -2\gamma [t_1 \phi_{L,1} \quad \dots \quad t_K \phi_{L,K} \quad \dots \quad t_N \phi_{L,N}]$$

$$f_2 = -2\gamma [t_1 \phi_{U,1} \quad \dots \quad t_K \phi_{U,K} \quad \dots \quad t_N \phi_{U,N}]$$

$$f_3 = \begin{bmatrix} C & \dots & C \\ \underbrace{\hspace{10em}}_{4N} \end{bmatrix}$$

The matrix A in the inequality is presented as follows.

$$A = \begin{bmatrix} A_{11} & A_{12} & -I & O & O & O \\ A_{21} & A_{22} & O & -I & O & O \\ A_{31} & A_{32} & O & O & O & -I \\ A_{41} & A_{42} & O & O & -I & O \\ O & O & -I & O & O & O \\ O & O & O & -I & O & O \\ O & O & O & O & -I & O \\ O & O & O & O & O & -I \end{bmatrix} \quad (45)$$

where:

$$A_{11} = -0.5 \text{diag}(\phi_{L,1}, \dots, \phi_{L,k}, \dots, \phi_{L,N}) \in R^{N \times N}$$

$$A_{12} = -0.5 \text{diag}(\phi_{L,1}, \dots, \phi_{L,k}, \dots, \phi_{L,N}) \in R^{N \times N}$$

$$A_{21} = 0.5 \text{diag}(\phi_{L,1}, \dots, \phi_{L,k}, \dots, \phi_{L,N}) \in R^{N \times N}$$

$$A_{22} = 0.5 \text{diag}(\phi_{R,1}, \dots, \phi_{R,k}, \dots, \phi_{R,N}) \in R^{N \times N}$$

$$A_{23} = O \in R^{N \times N}, \quad A_{42} = O \in R^{N \times N}$$

$$A_{23} = \text{diag}(\phi_{R,1}, \dots, \phi_{R,k}, \dots, \phi_{R,N}) \in R^{N \times N}$$

$$A_{41} = -\text{diag}(\phi_{L,1}, \dots, \phi_{L,k}, \dots, \phi_{L,N}) \in R^{N \times N}$$

The vector b in the inequality of (41) is represented as follows.

$$b = [b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5 \quad b_6 \quad b_7 \quad b_8]^T \quad (46)$$

With its components b_1, b_2, \dots, b_8 given as follows.

$$b_1 = [-t_1 + \varepsilon \quad \dots \quad -t_k + \varepsilon \quad \dots \quad -t_N + \varepsilon]^T$$

$$b_2 = [t_1 + \varepsilon \quad \dots \quad t_k + \varepsilon \quad \dots \quad t_N + \varepsilon]^T$$

$$b_3 = [t_1 + \varepsilon \quad \dots \quad t_k + \varepsilon \quad \dots \quad t_N + \varepsilon]^T$$

$$b_4 = [-t_1 + \varepsilon \quad \dots \quad -t_k + \varepsilon \quad \dots \quad -t_N + \varepsilon]^T$$

$$b_5, b_6, b_7, b_8 = O \in R^{N \times N}.$$

VI. SIMULATION RESULTS

To analyze the capability and performance of the proposed algorithm in predicting an appropriate prediction interval, it is used for several existing datasets in literature including input/output data associated with static function approximation [53], and datasets gathered from time varying dynamic systems [54]. The prediction is done using the constrained cost function represented in (28)-(33) (see Section V-A). The accuracy of the predictions is compared with existing methods to show the superior performance of the proposed approach over the state-of-the-art methods in literature. Three sets of parameter values are considered for the proposed algorithm which are listed in Table I.

TABLE I
PARAMETER SETS CONSIDERED FOR THE PROPOSED ALGORITHM IN THE FOLLOWING EXPERIMENTS

Parameter set	RMSE for test data
No. 1	$\varepsilon = 0.01, C = 40$ and $\gamma = 0.1$
No. 2	$\varepsilon = 0.01, C = 40$, and $\gamma = 1$
No. 3	$\varepsilon = 0.1, C = 40$, and $\gamma = 1$

A. Nonlinear Function Approximation

A nonlinear function approximation is considered as the first example. This nonlinear function has previously been considered in a number of papers [53] as follows:

$$y = 1 + x_1^{0.5} + x_2^{-1} + x_3^{-1.5} \quad (47)$$

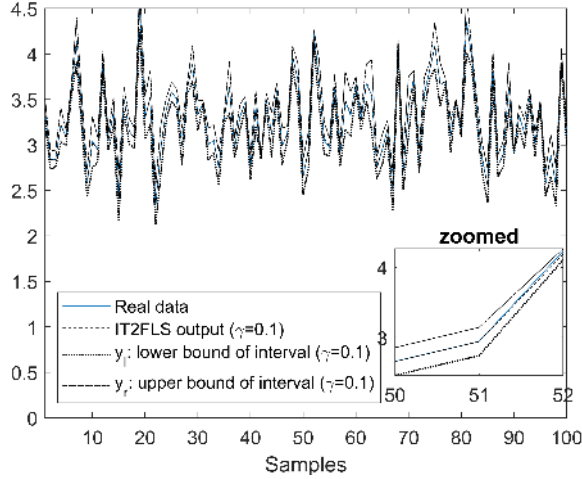


Fig. 3. Prediction Interval obtained for the nonlinear function approximation

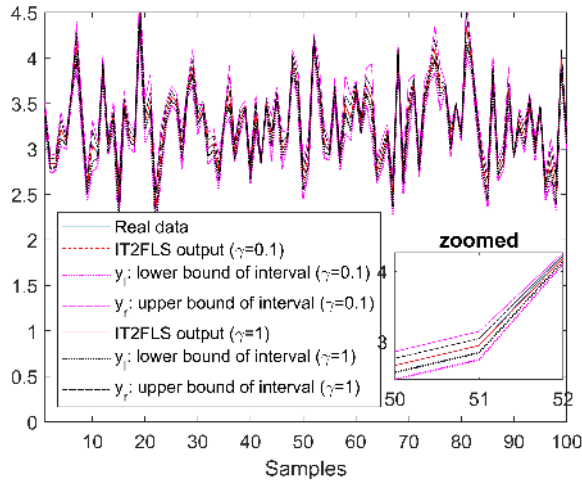


Fig. 4. Prediction Interval obtained for the nonlinear function approximation using different values of γ

To estimate this function 1000 randomly generated numbers are selected from the interval of [1, 5]. This data is further split into test and train data with 20% of points being considered for testing, and the rest being used for training purposes. MFs used in this study are Gaussian interval type-2 fuzzy MFs with crisp centre and interval σ values. Table II demonstrates the results of comparison between RMSEs of the proposed approach versus five other approaches previously investigated in [53,55]. All competitors listed in this table benefit from adaptive neuro-fuzzy inference systems constructed upon type-1 MFs with various training methods. As shown in table II, the proposed approach considerably outperforms other methods

TABLE II

IDENTIFICATION RESULTS FOR THE NONLINEAR FUNCTION APPROXIMATION PRESENTED IN EXAMPLE 2. BOLD FACED RESULTS INDICATES THE BEST ONE.

Method	RMSE for test data
ANFIS GD-RLSE [53]	0.245
ANFIS RPROP+RLSE [53]	0.218
ANFIS QP+RLSE [53]	0.773
ANFIS LM+RLSE [53]	0.288
ANFIS AWPSO+RLS [55]	0.072
Proposed approach with parameter set No. 1	0.016
Proposed approach with parameter set No. 2	0.017

in terms of generalization for test datasets. Furthermore, Fig. 3 illustrates that the prediction interval covers the target data samples. The capability of the proposed algorithms in controlling the prediction interval using the parameter γ is illustrated in Fig. 4; where the prediction interval for the case when γ is equal to 0.1 and 1 are demonstrated. It can be observed in Fig. 4 that a large value for γ results in a narrower prediction interval, such a result complies with the claimed role of γ in the cost function and estimation procedure.

B. Identification of a Time-Varying Nonlinear Dynamic System

A second order nonlinear dynamic system with time-varying parameters [54] is used to test the performance of the proposed approach in this study. The output of this dynamic system is a nonlinear time-varying function of inputs, time delays of input and time delayed output values as follows [54]:

$$y(t+1) = f(y(t), y(t-1), y(t-2), u(t), u(t-1)) \quad (48)$$

where the nonlinear function $f(\cdot)$ is defined as follows.

$$f(x_1, x_2, x_3, x_4, x_5) = \frac{x_1 x_2 x_3 x_5 (x_3 - b) + c x_4}{a + x_2^2 + x_3^2} \quad (49)$$

and parameters a , b and c are time-varying parameters defined as follows.

$$a(t) = 1.2 - 0.2 \cos\left(\frac{2\pi t}{T}\right)$$

$$b(t) = 1.0 - 0.4 \sin\left(\frac{2\pi t}{T}\right)$$

$$c(t) = 1.0 + 0.4 \sin\left(\frac{2\pi t}{T}\right)$$

with T , the total number of samples, is taken as to be equal to 1000. The input signal to the system $u(t)$ is taken as follows.

$$u(t) = \begin{cases} \sin\left(\frac{\pi t}{25}\right) & t < 250 \\ 1.0 & 250 \leq t < 500 \\ -1.0 & 500 \leq t < 750 \\ f(t) & 750 \leq t < 1000 \end{cases} \quad (50)$$

where:

$$f(t) = 0.3 \sin\left(\frac{\pi t}{25}\right) + 0.1 \sin\left(\frac{\pi t}{32}\right) + 0.6 \sin\left(\frac{\pi t}{10}\right).$$

TABLE III

IDENTIFICATION PERFORMANCE FOR THE TIME-VARYING SYSTEM IDENTIFICATION PROBLEM PRESENTED IN THIS PAPER. BOLD FACED RESULTS INDICATES THE BEST ONE.

Method	Rules	Epoch	Training RMSE	Testing RMSE
Type-1 TSK FNS [56]	9	100	0.0282	0.0598
Type-2 TSK FNS [56]	4	100	0.0284	0.0601
Feedorward Type-2 FNN	3	100	0.0281	0.0593
SIT2FNN [57]	4	100	0.0351	0.0560
SEIT2 FNN [58]	3	100	0.0274	0.0574
TSCIT2FNN [59]	3	100	0.0279	0.0576
IT2 FNN-GD [54]	-	200	0.0540	0.0613
IT2 FNN-SMC [54]	-	200	0.0360	0.0390
IT2 FNNPSO + SMC [54]	-	200	0.0199	0.0390
Proposed Approach with parameter set No. 3	11	Non-iterative	0.0146	0.0348

The first 80% of the generated data is used for training and the last 20% is chosen for testing purposes. Parameter values used for this system identification case are $\varepsilon = 0.1$, $\gamma = 1$, and $C = 40$.

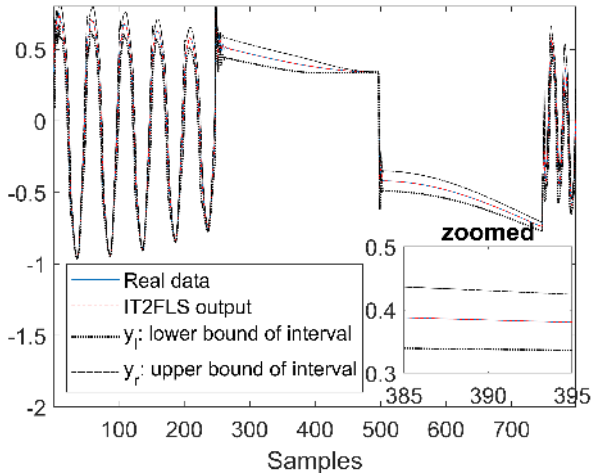


Fig. 5. Prediction performance of the proposed approach on time varying system identification in terms of the training dataset

The proposed algorithm is a non-iterative approach and does not include any iterations before convergence to true values of the parameters. A summary of the results obtained using the proposed method and several existing system identification models is given in Table III. As Table III shows, the proposed algorithm outperforms existing methods in terms of system identification accuracy for the training dataset as well as generalization to the test data. The identification performance is illustrated in Fig. 5 which shows that the output of the IT2FLS closely replicates the real data for training data. Furthermore,

TABLE IV

STATISTICS ASSOCIATED WITH THE URMIA LAKE WATER LEVEL DATASET.

Parameter	Value
Max.	6.5630
Min	-1.9430
Std.	2.4125
Mean	2.1689

TABLE V

PREDICTION RESULTS OVER THE URMIA LAKE WATER LEVEL DATASET. BOLD FACED RESULTS INDICATES THE BEST ONE.

	RMSE test	RMSE train
Trained ANFIS with hybrid (27 rules) [26]	0.3633	–
Trained ANFIS with PSO (27 rules) [26]	0.2454	0.43408
Hybrid PSO-RLS (27 rules) [26]	0.11089	0.224
Hybrid PSO-GD (27 rules) [26]	0.12621	0.239
ANFIS (1296 rules) [60]	0.08	–
The proposed method with with parameter set No. 3	0.0348	0.0146

the data is bounded with y_l and y_r , which provide an appropriate prediction interval for data. As mentioned earlier, such prediction intervals are controllable using appropriate selection of the parameter γ . It can be further shown by simulation that this prediction interval is never violated by desired output values which is a very appreciable result. In another words, the y_l 's never exceed corresponding desired output values and y_r 's are never smaller than the corresponding desired output values.

C. Discussion

Overall results demonstrate higher accuracy in prediction with respect to state-of-the-art papers in literature. Prediction interval estimation is a major feature of the proposed approach over existing literature which relaxes some of the assumptions required by previous approaches such as Gaussian probability distributed functions for prediction error. The prediction interval obtained using this approach is a reasonably inclusive one which is not too wide to avoid less specific information associated with data. The role of the parameter γ is also crucial in the performance of the proposed approach. The larger value for the parameter γ results in smaller width for prediction interval which contains more specific information about data. However, larger value for γ may result in having more points violating the prediction interval which needs to be avoided. Hence, the appropriate selection for the parameter γ is required for appropriate estimation of the prediction interval.

VII. APPLICATION TO PREDICTION INTERVAL IDENTIFICATION ASSOCIATED WITH URMIA LAKE WATER LEVEL USING SATELLITE REMOTE SENSING

Remote earth observation through satellites is crucial to gain valuable information about environmental conditions, lake water levels, and other vital factors. Such data analysis can be used to manage drinkable water to prevent disasters in the area [20]. Remote sensing is a major branch of earth

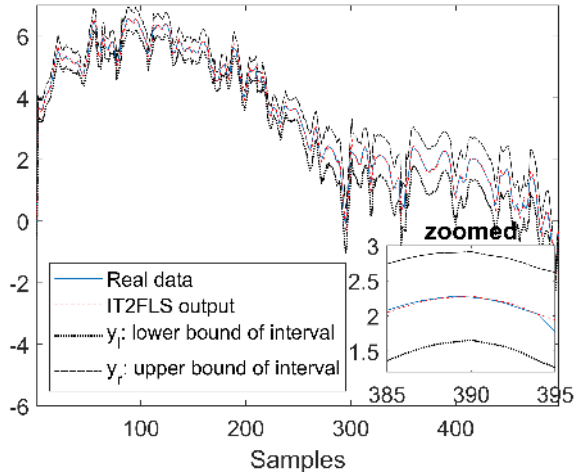


Fig. 6. Performance of the proposed approach on lake water level prediction for the training data set

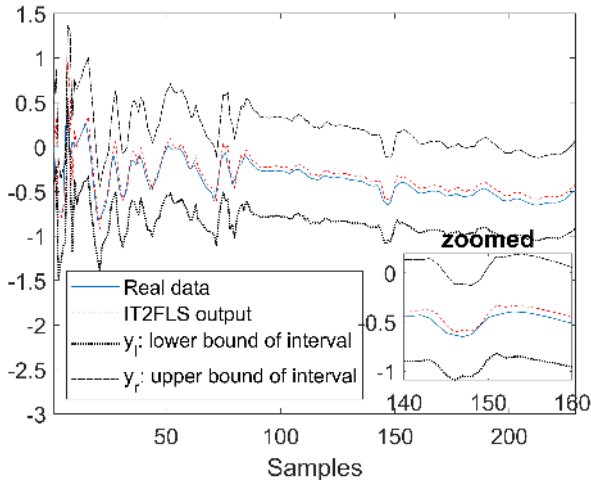


Fig. 7. Performance of the proposed approach on lake water level prediction for the training data set

observation science which provides a huge amount of information using satellite. A huge amount of data gathered from the environment require appropriate analysis to be useful for further analysis and decision making [61]. Machine learning approaches [62] as well as time series analysis [21] may be used to analyze collected data. Urmia lake is located in the span of longitudes of 45° to 46° east and latitudes of 37° to 38.5° north in the northwest of Iran. Fluctuation in the water level of lake Urmia has been paid attention due to its environmental and social impact on living conditions in nearby area [26]. The data associated with lake water level was collected with a sample interval of 10 days between 1992 and 2014 using satellites and includes 727 data samples, from which 509 samples are selected for training and 218 data samples are used for testing data. The collection of data was performed using the TOPEX/Poseidon/Jason satellite series

(at 10-day resolution) ². The statistics associated with the lake Urmia dataset water level are presented in Table IV. This dataset has previously been investigated in a number of papers in which adaptive neuro-fuzzy inference systems with various training algorithms were used to predict it [26,60]. The inputs taken for the prediction system in this study are $[y(t-40) \ y(t-30) \ y(t-10) \ y(t)]$ and the desired value of the system is $y(t+10)$ where $y(t-n)$ represents the lake water level n days ago. The results obtained using the proposed approach as well as several approaches in literature are presented in Table V.

As shown in Table V, the obtained values for the proposed algorithm outperform other approaches considerably while the number of rules considered for the proposed method is 5 which is considerably less than that of [60], [26]. The performance of the proposed approach as well as test data is indicated in Figs 6 and 7. As can be seen from these figures, the obtained results are close to measured lake water level.

VIII. CONCLUSIONS

In this study, the identification of the prediction interval associated with data is investigated. Three main categories of prediction interval estimation are distinguished within literature. The first category, explained in Section II-A, involves statistical approaches that perform uncertainty analysis on the model output. Using the co-variance associated with the model output as well as the properties of error with the normal probability distribution function, they find the confidence interval that covers future values of the output at certain level of confidence. The second class, described in Section II-B, involves the estimation of lower and upper bounds for data *a priori*, which is then used as a guideline to estimate model parameters. The third category, described in Section II-C, addresses the shortcomings of the first two categories including the assumption of the normal probability distribution function for error as well as finding the lower and upper bound of data by defining an appropriate cost function on data directly.

The method which is used to identify the prediction interval using IT2FLSs proposed in this study falls within the third category of estimation algorithms. Using the proposed approach, the prediction interval related to data using IT2FLSs is estimated with a multi-objective cost function. It is assumed that data values are not interval in nature, however, other than identification of their crisp values, it is desired to find a prediction interval covering the data. Using the multi-objective cost function, not only can we predict data, but we can also have a parameter to control the width of the prediction interval to prevent it from being unnecessarily large. Although the method investigated in this study falls into the third category of prediction intervals, having control over the prediction interval width is the most important contribution of this study with respect to the current approaches in this category. Benchmark datasets investigated in this study are

²Lake Urmia (0115) Height Variations from TOPEX POSEIDON and Jason series Altimetry

generated using a nonlinear function and a time-varying dynamic system. A quadratic programming method is used to minimize the constrained optimization problem and estimate the parameters of the IT2FLS. It is shown through simulation that with an appropriate choice of parameters associated with the proposed algorithm, it is possible to obtain a narrow yet covering prediction interval whilst maintaining the prediction accuracy. Comparisons between the proposed approach and state-of-the-art prediction methods in literature demonstrate the superior prediction accuracy of the proposed approach over them. It is further shown that using the proposed approach, an appropriate prediction interval can be estimated to cover data with reasonable number of rules without losing accuracy.

Motivated by the results obtained from the proposed approach over benchmark prediction methods, it is used to predict the Urmia lake water level collected through satellites. Such predictions may be used to predict a possible water shortage in the area and prevent it from happening. Simulation results demonstrated that the proposed approach is an effective way to deal with this time series dataset. Comparison results demonstrated superior performance of the proposed approach over the state-of-the-art approaches in literature. The prediction interval estimated in this method covers data and can be used for decision making purposes.

The limitation of the proposed approach is that it is a non-iterative approach which require the whole dataset in batch form to be used for parameter estimation. As future work, it would be interesting to study the recursive and possibly less computationally expensive version of this algorithm.

IX. ACKNOWLEDGEMENT

This work is funded and supported by the Engineering and Physical Sciences Research Council (EPSRC) under grant number: EP/R021031/1 - New Industrial Systems: Chatty Factories. Authors gratefully acknowledge Prof. Robert John for his comments and suggestion which improved the quality of this paper. Authors acknowledge Dr. Thomas Smith for his comments to improve the quality of this paper as well.

REFERENCES

- [1] M. A. Khanesar, E. Kayacan, M. Teshnehlab, and O. Kaynak, "Analysis of the noise reduction property of type-2 fuzzy logic systems using a novel type-2 membership function," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 5, pp. 1395–1406, 2011.
- [2] J. M. Mendel, "Uncertain rule-based fuzzy systems," in *Introduction and new directions*. Springer, 2017, p. 684.
- [3] J. A. Farrell and M. M. Polycarpou, *Adaptive approximation based control: unifying neural, fuzzy and traditional adaptive approximation approaches*. John Wiley & Sons, 2006, vol. 48.
- [4] I. Škrjanc, S. Blažič, and O. Agamennoni, "Identification of dynamical systems with a robust interval fuzzy model," *Automatica*, vol. 41, no. 2, pp. 327–332, 2005.
- [5] I. Škrjanc, S. Blazic, and O. Agamennoni, "Interval fuzzy model identification using L_{∞} -norm," *IEEE Transactions on fuzzy systems*, vol. 13, no. 5, pp. 561–568, 2005.
- [6] S. Oblak, "Interval fuzzy modelling in fault detection for a class of processes with interval-type parameters," in *EUROCON 2005-The International Conference on "Computer as a Tool"*, vol. 2. IEEE, 2005, pp. 1867–1870.
- [7] S. Oblak, I. Škrjanc, and S. Blazic, "On applying interval fuzzy model to fault detection and isolation for nonlinear input-output systems with uncertain parameters," in *Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005*. IEEE, 2005, pp. 465–470.
- [8] F. Valencia, J. Collado, D. Sáez, and L. G. Marín, "Robust energy management system for a microgrid based on a fuzzy prediction interval model," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1486–1494, 2015.
- [9] M. F. Zarandi, B. Rezaee, I. Turksen, and E. Neshat, "A type-2 fuzzy rule-based expert system model for stock price analysis," *Expert Systems with Applications*, vol. 36, no. 1, pp. 139–154, 2009.
- [10] S. Hassan, M. A. Khanesar, E. Kayacan, J. Jaafar, and A. Khosravi, "Optimal design of adaptive type-2 neuro-fuzzy systems: A review," *Applied Soft Computing*, vol. 44, pp. 134–143, 2016.
- [11] M. A. Khanesar, S. Hassan, E. Cambria, and E. Kayacan, "A novel non-iterative parameter estimation method for interval type-2 fuzzy neural networks based on a dynamic cost function," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2019, pp. 1–6.
- [12] Z. Deng, K.-S. Choi, L. Cao, and S. Wang, "T2fela: type-2 fuzzy extreme learning algorithm for fast training of interval type-2 tsk fuzzy logic system," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 4, pp. 664–676, 2013.
- [13] V. Uslan, H. Seker, and R. John, "A support vector-based interval type-2 fuzzy system," in *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2014, pp. 2396–2401.
- [14] C.-F. Juang, R.-B. Huang, and W.-Y. Cheng, "An interval type-2 fuzzy-neural network with support-vector regression for noisy regression problems," *IEEE Transactions on fuzzy systems*, vol. 18, no. 4, pp. 686–699, 2010.
- [15] L. Zhang and P. B. Luh, "Neural network-based market clearing price prediction and confidence interval estimation with an improved extended kalman filter method," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 59–66, 2005.
- [16] H. Papadopoulos and H. Haralambous, "Reliable prediction intervals with regression neural networks," *Neural Networks*, vol. 24, no. 8, pp. 842–851, 2011.
- [17] C. I. van Hinsbergen, J. Van Lint, and H. Van Zuylen, "Bayesian committee of neural networks to predict travel times with confidence intervals," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 5, pp. 498–509, 2009.
- [18] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Transactions on neural networks*, vol. 22, no. 9, pp. 1341–1356, 2011.
- [19] A. Núñez and B. De Schutter, "Distributed identification of fuzzy confidence intervals for traffic measurements," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 6995–7000.
- [20] X. Li, Q. Liu, R. Yang, J. Wen, J. Zhang, E. Cai, and H. Zhang, "The combination of ground-sensing network and satellite remote sensing in huailai county," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3819–3826, 2016.
- [21] B. R. Tripathy, V. Tiwari, V. Pandey, C. D. Elvidge, J. Rawat, M. Sharma, R. Prwasi, and P. Kumar, "Estimation of urban population dynamics using dmsp-ols night-time lights time series sensors data," *IEEE Sensors Journal*, vol. 17, no. 4, pp. 1013–1020, 2016.
- [22] X. Sun, C. Li, L. Liu, J. Yin, Y. Lei, and J. Zhao, "Dynamic monitoring of haze pollution using satellite remote sensing," *IEEE Sensors Journal*, 2019.
- [23] S. Hemalatha and S. M. Anuncia, "Unsupervised segmentation of remote sensing images using fd based texture analysis model and isodata," *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 8, no. 3, pp. 58–75, 2017.
- [24] S. Borra, R. Thanki, and N. Dey, *Satellite image analysis: Clustering and classification*. Springer, 2019.
- [25] N. Dey, C. Bhatt, and A. S. Ashour, "Big data for remote sensing: Visualization, analysis and interpretation," *Cham: Springer*, p. 104, 2018.
- [26] M. Esbati, M. A. Khanesar, and A. Shahzadi, "Modeling level change in lake urmia using hybrid artificial intelligence approaches," *Theoretical and Applied Climatology*, vol. 133, no. 1-2, pp. 447–458, 2018.
- [27] "lake rmial (0115) height variations from topex/poseidon and jason series altimetry." [Online]. Available: https://lipad.fas.usda.gov/cropexplorer/global_reservoir/gr_regional_chart_json1.aspx?regionid=metu&reservoir_name=Urmia-1
- [28] J. R. Magnus, O. Powell, and P. Prüfer, "A comparison of two model averaging techniques with an application to growth empirics," *Journal of econometrics*, vol. 154, no. 2, pp. 139–153, 2010.
- [29] A. Khosravi, S. Nahavandi, and D. Creighton, "A neural network-garch-based method for construction of prediction intervals," *Electric Power Systems Research*, vol. 96, pp. 185–193, 2013.

- [30] R. Dybowski and S. J. Roberts, "Confidence intervals and prediction intervals for feed-forward neural networks," *Clinical applications of artificial neural networks*, pp. 298–326, 2001.
- [31] I. Rivals and L. Personnaz, "Construction of confidence intervals for neural networks based on least squares estimation," *Neural Networks*, vol. 13, no. 4–5, pp. 463–484, 2000.
- [32] J. de Jesús Rubio, "Stable kalman filter and neural network for the chaotic systems identification," *Journal of the Franklin Institute*, vol. 354, no. 16, pp. 7444–7462, 2017.
- [33] M. A. Khameneh, E. Kayacan, M. Teshnehlab, and O. Kaynak, "Extended kalman filter based learning algorithm for type-2 fuzzy logic systems and its experimental evaluation," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4443–4455, 2011.
- [34] C. Guan, P. B. Luh, L. D. Michel, and Z. Chi, "Hybrid kalman filters for very short-term load forecasting and prediction interval estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3806–3817, 2013.
- [35] M. Z. Khan and M. F. Khan, "Application of anfis, ann and fuzzy time series models to co2 emission from the energy sector and global temperature increase," *International Journal of Climate Change Strategies and Management*, 2019.
- [36] Z. Pezeshki and S. M. Mazinani, "Comparison of artificial neural networks, fuzzy logic and neuro fuzzy for predicting optimization of building thermal consumption: a survey," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 495–525, 2019.
- [37] I. Škrjanc, "Fuzzy confidence interval for ph titration curve," *Applied mathematical modelling*, vol. 35, no. 8, pp. 4083–4090, 2011.
- [38] W. Li, J. Zhou, L. Chen, K. Feng, H. Zhang, C. Meng, and N. Sun, "Upper and lower bound interval forecasting methodology based on ideal boundary and multiple linear regression models," *Water Resources Management*, vol. 33, no. 3, pp. 1203–1215, 2019.
- [39] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE transactions on neural networks*, vol. 22, no. 3, pp. 337–346, 2010.
- [40] A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "A new fuzzy-based combined prediction interval for wind power forecasting," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 18–26, 2015.
- [41] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using pso-tuned support vector machines," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 2, pp. 322–331, 2015.
- [42] L. Ye, J. Zhou, H. V. Gupta, H. Zhang, X. Zeng, and L. Chen, "Efficient estimation of flood forecast prediction intervals via single-and multi-objective versions of the lube method," *Hydrological Processes*, vol. 30, no. 15, pp. 2703–2716, 2016.
- [43] N. A. Shrivastava, K. Lohia, and B. K. Panigrahi, "A multiobjective framework for wind speed prediction interval forecasts," *Renewable Energy*, vol. 87, pp. 903–910, 2016.
- [44] A. Kavousi-Fard, "Modeling uncertainty in tidal current forecast using prediction interval-based svr," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 708–715, 2016.
- [45] I. M. Galván, J. M. Valls, A. Cervantes, and R. Aler, "Multi-objective evolutionary optimization of prediction intervals for solar energy forecasting with neural networks," *Information Sciences*, vol. 418, pp. 363–382, 2017.
- [46] D. Senthilkumar and C. Mahanta, "Identification of uncertain nonlinear systems for robust fuzzy control," *ISA transactions*, vol. 49, no. 1, pp. 27–38, 2010.
- [47] X. Liu, Z. Yun, H. Yang, and Q. Zhang, "Fault detection based on lp-svr interval regression model with l1-norm minimization," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 3, pp. 3991–4001, 2019.
- [48] D. Wu, "Approaches for reducing the computational cost of interval type-2 fuzzy logic systems: overview and comparisons," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 1, pp. 80–99, 2012.
- [49] M. A. Khameneh and J. M. Mendel, "Maclaurin series expansion complexity-reduced center of sets type-reduction+ defuzzification for interval type-2 fuzzy systems," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2016, pp. 1224–1231.
- [50] D. Wu and J. M. Mendel, "Enhanced karnik-mendel algorithms," *IEEE transactions on fuzzy systems*, vol. 17, no. 4, pp. 923–934, 2008.
- [51] M. Biglarbegian, W. W. Melek, and J. M. Mendel, "On the stability of interval type-2 tsf fuzzy logic control systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 3, pp. 798–818, 2009.
- [52] M. Nie and W. W. Tan, "Towards an efficient type-reduction method for interval type-2 fuzzy logic systems," in *2008 IEEE international conference on fuzzy systems (IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 1425–1432.
- [53] M.-S. Chen, "A comparative study of learning methods in tuning parameters of fuzzy membership functions," in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, vol. 3. IEEE, 1999, pp. 40–44.
- [54] E. Kayacan and M. A. Khameneh, *Fuzzy neural networks for real time control applications: concepts, modeling and algorithms for fast learning*. Butterworth-Heinemann, 2015.
- [55] M. A. Shoorehdeli, M. Teshnehlab, and A. Sedigh, "Novel hybrid learning algorithms for tuning anfis parameters using adaptive weighted pso," in *2007 IEEE International Fuzzy Systems Conference*. IEEE, 2007, pp. 1–6.
- [56] R. H. Abiyev and O. Kaynak, "Type 2 fuzzy neural structure for identification and control of time-varying plants," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 12, pp. 4147–4159, 2010.
- [57] Y.-Y. Lin, S.-H. Liao, J.-Y. Chang, and C.-T. Lin, "Simplified interval type-2 fuzzy neural networks," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 5, pp. 959–969, 2013.
- [58] C.-F. Juang and Y.-W. Tsao, "A self-evolving interval type-2 fuzzy neural network with online structure and parameter learning," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1411–1424, 2008.
- [59] Y.-Y. Lin, J.-Y. Chang, and C.-T. Lin, "A tsf-type-based self-evolving compensatory interval type-2 fuzzy neural network (tsct2fnn) and its applications," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 1, pp. 447–459, 2013.
- [60] M. Talebizadeh and A. Moridnejad, "Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ann and anfis models," *Expert Systems with applications*, vol. 38, no. 4, pp. 4126–4135, 2011.
- [61] Y. Jiang, S. Yin, J. Dong, and O. Kaynak, "A review on soft sensors for monitoring, control and optimization of industrial processes," *IEEE Sensors Journal*, 2020.
- [62] N. Zerrouki, F. Harrou, Y. Sun, and L. Hocini, "A machine learning-based approach for land cover change detection using remote sensing and radiometric measurements," *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5843–5850, 2019.



M. A. Khameneh is currently working as a post-doctoral research fellow within Advanced Manufacturing Technology Research Group at the University of Nottingham, UK. He has further held previous positions as a postdoctoral researcher at the Technical University of Denmark and an assistant professor at Semnan University, Iran. He is also a senior member of IEEE and serves as an academic editor for Complexity journal, a publication of John Wiley & Sons in collaboration with Hindawi Publishing Corporation. He received B.Sc., M.S., and Ph.D. degrees in control engineering from K. N. Toosi University of Technology, Tehran, Iran, in 2005, 2007, and 2012, respectively. In 2010, he held a nine-month visiting student position at the Bogazici University, Istanbul, Turkey. His current research interests include manufacturing, robotic, machine learning and control.



David T. Branson, III received the undergraduate and master's degrees in mechanical engineering from the University of Wisconsin—Madison, Madison, WI, USA, in 1999 and 2002, respectively, and the Ph.D. degree in non-linear modelling and control of multi-body systems from the University of Bath, Bath, U.K., in 2006. He is currently an Associate Professor with the Faculty of Engineering, University of Nottingham, Nottingham, U.K., and the Director of the Nottingham Advanced Robotics Laboratory. His

research interests include modeling and control of nonlinear systems undergoing large deformations/displacements with application to the design and control of robotic systems for manufacturing and health-care purposes, wearable monitoring systems, and controlled continuum arms/surfaces.