Statistically Evolving Fuzzy Inference System for Non-Gaussian Noises

Zhao-Xu Yang, Hai-Jun Rong*, Plamen Angelov, Zhi-Xin Yang*

Abstract—Non-Gaussian noises always exist in the nonlinear system, which usually lead to inconsistency and divergence of the regression and identification applications. The conventional evolving fuzzy systems (EFSs) in common sense have succeeded to conquer the uncertainties and external disturbance employing the specific variable structure characteristic. However, non-Gaussian noises would trigger the frequent changes of structure under the transient criteria, which severely degrades performance. Statistical criterion provides an informed choice of the strategies of the structure evolution, utilizing the approximation uncertainty as the observation of model sufficiency. The approximation uncertainty can be always decomposed into model uncertainty term and noise term, and is suitable for the non-Gaussian noise condition, especially relaxing the traditional Gaussian assumption. In this paper, a novel incremental statistical evolving fuzzy inference system (SEFIS) is proposed, which has the capacity of updating the system parameters, and evolving the structure components to integrate new knowledge in the new process characteristic, system behavior, and operating conditions with non-Gaussian noises. The system generates a new rule based on the statistical model sufficiency which gives so insight into whether models are reliable and their approximations can be trusted. The nearest rule presents the inactive rule under the current data stream and further would be deleted without losing any information and accuracy of the subsequent trained models when the model sufficiency is satisfied. In our work, an adaptive maximum correntropy extend Kalman filter (AMCEKF) is derived to update the parameters of the evolving rules to cope with the non-Gaussian noises problems to further improve the robustness of parameter updating process. The parameter updating process shares an estimate of the uncertainty with the criteria of the structure evolving process to make the computation less of a burden dramatically. The simulation studies show that the proposed SEFIS has faster learning speed and is more accurate than the existing evolving fuzzy systems (EFSs) in the case of noise-free and noisy conditions.

Index Terms—Evolving fuzzy system, maximum correntropy, Kalman filter, model sufficiency.

I. Introduction

Numerous evolving fuzzy systems (EFSs) with remarkable flexibility, linguistic interoperability and dynamical autonomy

Zhao-Xu Yang is with State Key Laboratory for Strength and Vibration of Mechanical Structures, Shaanxi Key Laboratory of Environment and Control for Flight Vehicle, School of Aerospace Engineering, Xi'an Jiaotong University, P.R.China, and State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao. Hai-Jun Rong is with State Key Laboratory for Strength and Vibration of Mechanical Structures, Shaanxi Key Laboratory of Environment and Control for Flight Vehicle, School of Aerospace Engineering, Xi'an Jiaotong University, P.R.China. Plamen Angelov is with School of Computing and Communications, Lancaster University, United Kingdom. Zhi-Xin Yang is with Department of Electromechanical Engineering, State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao. *Corresponding author: hjrong@xjtu.edu.cn, zxyang@um.edu.mo

were introduced during the last two decades, which could be seen as a synergy between fuzzy systems as evolvable structures for information representation and computationally efficient method for lifelong self-learning to address non-stationary process [1, 2]. Consequentially, the mechanisms for structure evolution and parameter adaption are crucial for the overall performance of online learning in the case of significant dynamic system changes and non-stationary environments, which has motivated the development of various application domains, such as regression [3], identification [4], and classification [5].

Different criteria were used to trigger the structural evolution process. In [6–8], the self-organizing fuzzy neural network (SOFNN), the self-evolving fuzzy system (SEFS) and the metacognitive neuro-fuzzy inference system (McFIS) adopted the error criterion between the estimated and the real value for rule generation. However, the error criterion is very susceptible to the outliers, which would cause frequent changes. Moreover, when the error of the system remains high within a time range under an inappropriate threshold, the system is forced to grow the rules violently using the error criterion. Consequentially, the system obtains too many rules obviously and achieves the short-term Band-Aid solution, which is also known as the issue of overfitting. In [9–13], an Euclidean distance-based strategy between the new data and all the existing rules in the dynamic evolving neural-fuzzy inference system (DENFIS), the self-organizing fuzzy modified least-square (SOFMLS), the flexible fuzzy inference system (FLEXFIS), and the evolving fuzzy participatory learning (ePL) was utilized to implement a scatter partitioning of the input space to create fuzzy inference rules. Euclidean distance was used when eliciting the winner cluster, triggering ellipsoidal clusters parallel to the axes, while the Mahalanobis distance measure to the clustering process would trigger ellipsoidal clusters in arbitrary position. In [14], evolving possibilistic fuzzy modelling (ePFM) employed participatory learning to adapt the model structure using a Mahalanobis distance criterion. The distance criterion provided the efficient approach to evaluate the novelty of the datum. Moreover, a kind of evolving Takagi-Sugeno models (eTSs) [15, 16] were proposed that employs the recursive calculation of the information potential of the new data sample, which represents the spatial proximity measure used to define the antecedent parameters.

Except the instantaneous information of each datum, the statistical contribution of the rules defined as the significance criterion was proposed as an effective way to determine the structural strategy. The sequential adaptive fuzzy inference system (SAFIS) [17] and its improved vision of Extended SAFIS (ESAFIS) [18] used the concept of influence of a fuzzy rule to the overall output based on all input datum, in addition to the distance criterion. In [19], a parsimonious network

based on fuzzy inference system (PANFIS) with the multidimensional membership functions employed an extended rule significance which is extended from the concept in SAFIS by integrating hyperplanes consequents and generalizing to ellipsoids in arbitrary position.

Excellent results in the existing EFSs has been illustrated adequately from multiple aspects to demonstrate the effectiveness of the criteria. However, the scale of global rules are lack of adequacy measurement. In the existing criteria, error criterion describes the transient performance of the overall system. Distance criterion and informative potential give the novelty measurement of the current datum, and the significance criterion is the impact degree of each local rule which influence the overall output. The general condition for recruiting new rules includes one or any combination of two criteria when the structure is considered under inadequately dynamically capturing capacity, however, the drawback which is inevitably introduced can be either susceptibility to the outliers or computational complexity. Besides, to cope with the non-Gaussian noises, such as impulsive noises with heavy-tailed distributions, which are common in many real scenarios of nonlinear identification and regression, the performance adopting these criteria would be significantly deteriorated as drifting with the data stream volatility and confusing by the outliers.

Considering the non-Gaussian noises, the learning algorithms which are highly efficient and have better learning performance for systems with various statistics are keenly sought. Due to the merits of correntropy in handling the non-Gaussian noises with strong outliers rejection ability, a correntropy-based evolving fuzzy neural system (CEFNS) [20] was proposed based on the correntropy concept of the information theoretic learning to achieve the evolution of the rules and adjustment of the rule parameters. The CEFNS relied on the gradient ascent method to update the system parameters under the maximum correntropy criterion, which generally suffered from the convergence problem that was determined by the learning step. A recursive maximum correntropy based evolving fuzzy systems (RMCEFS) was proposed [21], in which the analytical and recursive solution of the maximum correntropy cost function was utilized to update the consequent parameters of the recruited rules. The architected condition for recruiting rules was a combination of distance criterion and the proposed maximum correntropy criterion that came down in one continuous line from the CEFNS. In both CEFNS and RMCEFS, an exponential function of the error as the extra scaling factor was utilized to depict the outlier rejection property. Although the existing EFSs have achieved satisfied accuracy performance to reduce the effects of non-Gaussian noises with the distance criterion and the correntropy born of the error criterion, the congenital flaw of overmany parameters restricts the lightweight design and modest resource requirements. In this review, we mainly concentrate on how EFS can be constructed for non-Gaussian noises with better global convergence and less consumerism. A novel statistically evolving fuzzy inference system (SEFIS) is proposed, and the salient contributions of this paper are organized as follows.

1) We provide a uniform criterion for recruiting and pruning rules based on the statistical concept in minimalism, named as the statistical model sufficiency from a global perspective. The statistical criterion provides an informed choice of the strategies of the structure evolution to eliminate the disturbances from different noises. The approximation uncertainty as the observation of model sufficiency in the criterion is utilized which can be always decomposed into model uncertainty term and noise term. This is suitable for the non-Gaussian noise model, especially relaxing the traditional Gaussian assumption. The conventional criteria for the evolving system structure, are hard to link directly to model sufficiency. Besides, the existing recruiting process [8, 17] are generally consisted of several criteria, such as error criterion, distance criterion and so on, which increases the computation complexity. In this way, the utilization of the only model sufficiency is optimal either in the expressibility or in the computational burden.

2) The proposed statistical approximation uncertainty realizes the rule evolution by describing patterns of variability and covariation. It is represented by the system output and noise covariance. To achieve its computation, a novel adaptive maximum correntropy extended Kalman filter (AMCEKF) is proposed. In our work, another merit of the AMCEKF is to update the consequence parameters under the non-Gaussian noise circumstance. The AMCEKF is an extension of EKF in which the minimum mean square criterion is replaced by maximum correntropy criterion. In its strictest sense, the evolution criterion is independent of the developed parameter updating algorithm. There has been, nevertheless, a legitimately valuable computational links, where the parameter updating process has been sharing an estimate of the uncertainty with the structural evolution process. It has further made the computation less of a burden.

3) The two general thresholds in the statistical rule evolution criterion are dimensionless statistic. They represent the level of significance. However, the existing rule recruiting criteria [8] may require some thresholds and generally are obtained by trial and error in response to different applications. Thus, the threshold of statistical optimality in this work has the strong universality, which releases the workload of the parameter tuning.

The rest of this paper is organized as follows. Section II gives the problem foundation. Section III describes the statistical model sufficiency criterion used in the structural evolution process. Section IV presents the adaptive maximum correntropy extended Kalman filter used in the parameter updating process. Learning process of SEFIS is designed in Section V. The performance evaluation is illustrated in Section VI and Section VII gives the conclusions of this work.

II. Problem Foundation

A general class of MIMO nonlinear systems can be described by the following discrete model,

$$\mathbf{y}(k) = \mathbf{f}\big(\mathbf{y}(k-1), \mathbf{y}(k-2), \ldots, \mathbf{y}(k-q)\big) + \boldsymbol{\eta}_k \qquad (1)$$

where $\mathbf{y}(k) = [y_{1k}, \ldots, y_{N_y k}]^T$. $\mathbf{f}(\cdot)$ is an unknown nonlinear mapping function, k is time script, and p is the maximum

lags. The system is considered in the presence of the non-Gaussian noises that are expressed as an interference term η_k. The external impulsive noises are a kind of common non-Gaussian noises, which are of heavy-tailed distribution. Besides, the noises of the Cauchy distribution and uniform distribution also belong to non-Gaussian noises.

When the uncertainty caused by the negative interference from the exterior environment grows more global, Gaussian distribution can be insufficient and becomes too crude as an approximation to the true posterior belief. Thus, we have to look for an efficient way to get over the resulting failure in the non-Gaussian situations. This limitation has been well recognized, and a common and popular method is the Gaussian paradigm extensions using mixtures of Gaussians to represent multimodal beliefs, e.g. approximating the predictive distribution by weighted Gaussian mixtures [22, 23]. The considered external non-Gaussian disturbances can be expressed with a mixed-Gaussian distribution, such that,

$$\eta_k = \sum_{i=1}^{N} s_i \eta_{ki} \tag{2}$$

where, $s_i(\sum_{i=1}^{N} s_i = 1)$ is the weight of the different superimposed Gaussian noise η_{ki}, and as a special case, some of these consisted Gaussian noises with the smallest possible variance σ_i can be analogized as the impulsive noises. Obviously, the aforementioned mixed-Gaussian distribution is a more general description of many extremes of the external disturbance, either containing one single Gaussian noise indicated as $\eta_k \sim N(\mu, \sigma), N = 1$, or at multiple random values of (μ_i, σ_i) indicated as $\eta_k = \sum_{i=1}^{N} s_i \eta_{ki}, \eta_{ki} \sim N(\mu_i, \sigma_i)$. The proposed EFS is potential to solve more general noise situations, if it can work well under the typical extreme of the non-Gaussian noise.

Selecting $\{y(k-1), \ldots, y(k-q)\}$ as the mapping function's input $\mathbf{x}_k$, Eq.(1) can be rewritten as

$$\mathbf{y}(k) = \mathbf{f}(\mathbf{x}_k) + \eta_k \tag{3}$$

In our work, the SEFIS algorithm is utilized to approximate $\mathbf{f}(\cdot)$ and restrain the negative impacts of the non-Gaussian noises such that

$$\hat{\mathbf{y}}(k) = \hat{\mathbf{f}}(\mathbf{x}_k) \tag{4}$$

where, $\hat{\mathbf{y}}(k)$ is the output of SEFIS. This means that the objective is to minimize the error between the system output and the approximator output, $\|\mathbf{y}(k) - \hat{\mathbf{y}}(k)\|$ under the non-Gaussian noises. Before describing the details of the algorithm, the structure of SEFIS is first described below.

The SEFIS is comprised of Takagi-Sigeno (T-S) type rules that are constructed in an adaptive and evolvable way based on the given input-output data. The nth rule has the following form,

Rule n:if $(x_{1k}$ is $U_{1n}^k) \cdots$ and $(x_{ik}$ is $U_{in}^k) \cdots$ and $(x_{N_xk}$ is $U_{N_xn}^k)$

then $(\hat{y}_{1k}$ is $a_{1n}^k) \cdots$ and $(\hat{y}_{jk}$ is $a_{jn}^k) \cdots$ and $(\hat{y}_{N_yk}$ is $a_{N_yn}^k)$

where $a_{jn}^k(j = 1, \ldots, N_y, n = 1, \ldots, N_h)$ is a constant consequent parameter in nth rule. $U_{in}^k(i = 1, \ldots, N_x)$ is the fuzzy set. N_x is the dimension of the input vector $\mathbf{x}_k = [x_{1k}, \ldots, x_{N_xk}]^T$,

N_h is the number of the fuzzy rules, N_y is the dimension of the output vector $\hat{\mathbf{y}}(k) = [\hat{y}_{1k}, \ldots, \hat{y}_{N_yk}]^T$.

With the weighted average method, the approximation function of the output $\hat{\mathbf{y}}(k)$ is computed as

$$\hat{\mathbf{f}}(\mathbf{x}_k) = \sum_{n=1}^{N_h} \bar{R}_n(\mathbf{x}_k) \mathbf{a}_n^k \tag{5}$$

where, $\mathbf{a}_n^k = [a_{1n}^k, \ldots, a_{N_yn}^k]^T$, and $\bar{R}_n(\mathbf{x}_k)$ is the nth normalized firing strength which is given by

$$\bar{R}_n(\mathbf{x}_k) = \frac{R_n(\mathbf{x}_k)}{\sum_{n=1}^{N_h} R_n(\mathbf{x}_k)} \tag{6}$$

where, $R_n(\mathbf{x}_k)$ is nth firing strength and calculated using unidimensional Gaussian membership function with center μ_{in}^k and width σ_n^k,

$$R_n(\mathbf{x}_k) = \prod_{i=1}^{N_x} A_{in}^k(x_i) = \prod_{i=1}^{N_x} \exp\left(-\frac{(x_{ik}-\mu_{in}^k)^2}{(\sigma_n^k)^2}\right) = \exp\left(-\frac{\|\mathbf{x}_k-\boldsymbol{\mu}_n^k\|^2}{(\sigma_n^k)^2}\right) \tag{7}$$

where $\boldsymbol{\mu}_n^k = [\mu_{1n}^k, \ldots, \mu_{N_xn}^k]^T$. The membership degree A_{in}^k of the ith input variable x_{ik} in nth rule reflects the degree of satisfaction with the fuzzy set U_{in}^k.

Remark 1. *Illustrated by Eq.(7), the fuzzy membership function is applied to obtain the fuzzy membership for a fuzzy set. In our proposed SEFIS, there will be no restriction on the type of fuzzy membership functions. Other type of fuzzy membership functions, such as triangle and trapezoid, are also feasible. Based on these membership functions, the fuzzy system also can obtain good approximation capability. Considering Gaussian membership function is commonly used one with smooth and strictly positive definite properties, we choose it as the fuzzy membership function.*

The infrastructure of proposed SEFIS described in Eq.(5)$\sim$ Eq.(7) is similar with the other existing EFSs [8, 17]. Although their infrastructures are similar, the targeted problems are different. It is required to evolve system structure and update parameter by integrating new knowledge under the non-Gaussian noise environment. The common used transient criterion or the other supportive criterion are inefficient either in the reliability or in the computational consumption. One of the key problem in our work that has to be settled is screening the appropriate criterion to determine the strategies of the structural evolution to evade the unreasonable pruning and recruiting resulted by the non-Gaussian noises or outliers from the input data stream. The another problem between optimal parameters $\mathbf{a}_n^k$ and current moody data that contains knowledge concerns the strong outliers rejection ability as well as the convergence speed. In this paper, the two problems are realized based on the statistical concept and correntropy theory. They are described below in detail.

III. Statistical Model Sufficiency Criterion

Recruiting or pruning the rules to obtain the optimal structural scale is highly significant for EFSs, which would be triggered perceiving the change of the model sufficiency immediately. Under the condition of the non-Gaussian noises,

$$var(\mathbf{y}(N+1)|\mathbf{x}_{N+1})$$

$$=var\left(\mathbf{f}(\mathbf{x}_{N+1}) + \sum_{i=1}^{N} s_i \boldsymbol{\eta}_{(N+1)i}\right)$$

$$=\mathbb{E}\left[\left(\mathbf{f}(\mathbf{x}_{N+1}) + \sum_{i=1}^{N} s_i \boldsymbol{\eta}_{(N+1)i}\right)^2\right] - \left(\mathbb{E}\left[\mathbf{f}(\mathbf{x}_{N+1}) + \sum_{i=1}^{N} s_i \boldsymbol{\eta}_{(N+1)i}\right]\right)^2$$

$$=\mathbb{E}\left[(\mathbf{f}(\mathbf{x}_{N+1}))^2 + \sum_{i=1}^{N}(s_i \boldsymbol{\eta}_{(N+1)i})^2 + 2\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_{N+1})(s_i \boldsymbol{\eta}_{(N+1)i}) + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\left(s_i \boldsymbol{\eta}_{(N+1)i}\right)\left(s_j \boldsymbol{\eta}_{(N+1)j}\right)\right]$$

$$-\left((\mathbb{E}[\mathbf{f}(\mathbf{x}_{N+1})])^2 + \sum_{i=1}^{N}\left(\mathbb{E}\left[s_i \boldsymbol{\eta}_{(N+1)i}\right]\right)^2 + 2\sum_{i=1}^{N}\left(\mathbb{E}[\mathbf{f}(\mathbf{x}_{N+1})]\mathbb{E}\left[s_i \boldsymbol{\eta}_{(N+1)i}\right]\right) + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathbb{E}\left[s_i \boldsymbol{\eta}_{(N+1)i}\right]\right)\left(\mathbb{E}\left[s_j \boldsymbol{\eta}_{(N+1)j}\right]\right)\right) \quad (8)$$

$$=\left(\mathbb{E}\left[(\mathbf{f}(\mathbf{x}_{N+1}))^2\right] - (\mathbb{E}[\mathbf{f}(\mathbf{x}_{N+1})])^2\right) + \left(\sum_{i=1}^{N}\mathbb{E}\left[(s_i \boldsymbol{\eta}_{(N+1)i})^2\right] - \sum_{i=1}^{N}\left(\mathbb{E}\left[s_i \boldsymbol{\eta}_{(N+1)i}\right]\right)^2\right)$$

$$+ 2\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[\left(s_i \boldsymbol{\eta}_{(N+1)i}\right)\left(s_j \boldsymbol{\eta}_{(N+1)j}\right)\right] - \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathbb{E}\left[s_i \boldsymbol{\eta}_{(N+1)i}\right]\right)\left(\mathbb{E}\left[s_j \boldsymbol{\eta}_{(N+1)j}\right]\right)\right)$$

$$=var\left(\mathbf{f}(\mathbf{x}_{N+1})\right) + \sum_{i=1}^{N} var\left(s_i \boldsymbol{\eta}_{(N+1)i}\right) + 2\sum_{i=1}^{N}\sum_{j=1}^{N} cov\left(s_i \boldsymbol{\eta}_{(N+1)i}, s_j \boldsymbol{\eta}_{(N+1)j}\right)$$

---

it is not expected to be sensitive due to the outliers. A statistical criteria to judge the model sufficiency from a structural global perspective is imperative to replace the significance criterion based on each local rule. The quantification of the uncertainty in the approximation process is critical for assessing how much to trust the quality of the nonlinear learning approach, which has profound criteria in the detection of model sufficiency. We have denoted a fuzzy system as function $\hat{\mathbf{f}}(\mathbf{x}_k)$, which captures the model architecture. It is assumed that the model and each of the component $\boldsymbol{\eta}_{ki}$ of the mixed-Gaussian noises are independent events. For a trained EFS where the parameters has been estimated along with its error covariance matrix, the EFS output uncertainty can be determined.

Given a set of N observations $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, and $Y = \{\mathbf{y}(1), \ldots, \mathbf{y}(N)\}$. The fuzzy system aims at locating some local minimum $\boldsymbol{\theta}$ of the cost function in the approximation, to be used in inference on new data. While it is hard to predict how accurate these approximations will be in practice, their form at least gives some idea of the interaction between the posterior expected output of the given input, as well as the posterior expected regression variance, and the behavior of the system output for the parameter in the vicinity of the optimal[24]. In particular, the variance of the approximation distribution quantifies the approximation uncertainty with a new data $\mathbf{x}_{N+1}$, which can be decomposed using law of total variance as shown in Eq.(8).

Due to $\mathbb{E}\left[\mathbf{f}(\mathbf{x}_{N+1})\left(s_i \boldsymbol{\eta}_{(N+1)i}\right)\right] - \left(\mathbb{E}[\mathbf{f}(\mathbf{x}_{N+1})]\mathbb{E}\left[s_i \boldsymbol{\eta}_{(N+1)i}\right]\right) = cov\left(\mathbf{f}(\mathbf{x}_{N+1}), s_i \boldsymbol{\eta}_{(N+1)i}\right) = 0$, we can see that the variance in Eq.(8) is decomposed into two terms. The first term $var\left(\mathbf{f}(\mathbf{x}_{N+1})\right)$ can be represented as follows,

$$\sigma_y^2 = var(\hat{\mathbf{y}}) \quad (9)$$

which reflects our ignorance over model parameter $\boldsymbol{\theta}$, referred to as the model uncertainty. The second term is written as follows,

$$\sigma_{ed}^2 = \sum_{i=1}^{N} var\left(s_i \boldsymbol{\eta}_{(N+1)i}\right) + 2\sum_{i=1}^{N}\sum_{j=1}^{N} cov\left(s_i \boldsymbol{\eta}_{(N+1)i}, s_j \boldsymbol{\eta}_{(N+1)j}\right)$$
$$= cov\left(\sum_{i=1}^{N} s_i \boldsymbol{\eta}_{(N+1)i}\right) \quad (10)$$

which is the noise level during data generating process. Then, a reliable reference interval of model sufficiency for the error norm can be constructed by

$$\left[0, z_d \sqrt{\sigma_y^2 + \sigma_{ed}^2}\right] \quad (11)$$

where, z_d is the value of the T-statistic at $d\%$ level of significance, and $\sqrt{\sigma_y^2 + \sigma_{ed}^2}$ derived from the $var(\mathbf{y}(N+1)|\mathbf{x}_{N+1})$ in Eq.(8) reflects the uncertainty in the expected output for the given input observation which can be computed as part of the following parameter estimation algorithm and hence there is no computational overhead in testing for model sufficiency. The reference interval in Eq.(11) is critical for approximation tasks, which means the error norm lies within a bound with a certain level of confidence. Model insufficiency will be warned when the error norm falls outside the constructed interval. So the null hypothesis for the statistical inference of model sufficiency is expressed as follows,

$$\mathcal{H}_0 : \frac{\|\mathbf{e}_k\|}{\sqrt{\sigma_y^2 + \sigma_{ed}^2}} \le z_d \quad \text{for } d\% \text{ of data} \quad (12)$$

This implies the null hypothesis could mean that the approximation error has fairly high probability of locking in the constructed interval, but don't rule out the matter of events of vanishingly small probability unpredictably happening. $\mathbf{e}_k$ is the approximation error for the observation $(\mathbf{x}_k, \mathbf{y}(k))$ which is given by

$$\mathbf{e}_k = \mathbf{y}(k) - \hat{\mathbf{y}}(k) \quad (13)$$

If the condition for $\mathcal{H}_0$ is violated, the null hypothesis is rejected so that the SEFIS scale is insufficient. The rejection implies that the system scale must be increased to match the complexity of the unknown nonlinear functions $\mathbf{f}(\cdot)$, and hence a new rule is added to the SEFIS. The rule recruiting to be sufficient system can effectively increase the complexity of the fuzzy inference system when the scale beneficial results are low. Meanwhile, if the model sufficiency is enough to be satisfied, in other word, the approximation error falls inside the predefined minor constructed interval, it can be assumed the structural scale is efficient with some rule redundancy. Then the pruning strategy is done using an inactive rule concept where the nearest rule is defined as the one closest to the input data and further would be deleted without losing any information and accuracy of the subsequent trained models.

Remark 2. *The proposed model sufficiency Eq.(12) refers to the notion of system structure with appropriate quantities of rules, which is deemed to be sufficient if the approximation error is within a certain level of confidence exhibited by the approximation system. This feature will help to avoid recruiting the rules blindly with the action of the error. If the approximation uncertainty and external noises are high, the approximation system may achieve large errors which are within the expectations of the system and hence it is not necessary to add a new rule.*

IV. ADAPTIVE MAXIMUM CORRENTROPY EXTENDED KALMAN FILTER

A. Calculation of approximation uncertainty

From $\mathcal{H}_0$, one can see that the approximation uncertainty of the statistical properties requires calculating the covariance of output and noise in different time, which is brought into the Kalman filter naturally due to the indispensable and shared calculation process. The traditional Kalman filter works well under Gaussian noise, but its performance may deteriorate significantly under non-Gaussian noises. The main reason for this is that Kalman filter is developed based on the minimum mean square error criterion, which captures only the second order statistics of the error signal and is sensitive to large outliers. To address this problem, an adaptive maximum correntropy extend Kalman filter (AMCEKF) is proposed to calculate the covariance in different time. Simultaneously, the AMCEKF is presented to update the parameters based on the correntropy theory, which can perform better in non-Gaussian noise environments, since correntropy contains second and higher order moments of the error. This will be described below.

We can take the output $\hat{\mathbf{y}}(k)$ with the parameter vector $\boldsymbol{\theta}(k) = [\mathbf{a}_1^{kT}, \ldots, \mathbf{a}_{Nh}^{kT}]^T \in R^{N_h N_y \times 1}$ existing in all the fuzzy rules. It is analogous to a nonlinear system described by the following state and measurement equations

$$\boldsymbol{\theta}(k) = F(\boldsymbol{\theta}(k-1)) + \mathbf{q}(k-1) \tag{14a}$$

$$\mathbf{y}(k) = H(\boldsymbol{\theta}(k)) + \mathbf{r}(k) \tag{14b}$$

where, F and H denote the continuously differentiable non-linear state function and nonlinear measurement function,

respectively. $\mathbf{q}(k-1)$ and $\mathbf{r}(k)$ denote the process noise and the measurement noise, which are mutually uncorrected with zero means. The covariance matrices satisfy

$$cov(\mathbf{q}(k-1)) = \mathbf{Q}(k-1) \tag{15a}$$

$$cov(\mathbf{r}(k)) = cov\left(\sum_{i=1}^{N} s_i \boldsymbol{\eta}_{(N+1)i}\right) = \mathbf{R}(k) \tag{15b}$$

The prior mean and corresponding covariance matrix are given by

$$\hat{\boldsymbol{\theta}}(k|k-1) = F\left(\hat{\boldsymbol{\theta}}(k-1|k-1)\right) \tag{16a}$$

$$\mathbf{P}(k|k-1) = \mathbf{F}(k-1)\mathbf{P}(k-1|k-1)\mathbf{F}^T(k-1) + \mathbf{Q}(k-1) \tag{16b}$$

where, $\mathbf{F}(k-1)$ is the Jacobian matrix gradient given by

$$\mathbf{F}(k-1) = \left.\frac{\partial F}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(k-1|k-1)} \tag{17}$$

We approximate the measurement in Eq.(14) by

$$\mathbf{y}(k) \approx H\left(\hat{\boldsymbol{\theta}}(k|k-1)\right) + \mathbf{H}(k)\left(\boldsymbol{\theta}(k) - \hat{\boldsymbol{\theta}}(k|k-1)\right) + \mathbf{r}(k) \tag{18}$$

where, $\mathbf{H}(k)$ is the Jacobian matrix gradient given by,

$$\mathbf{H}(k) = \left.\frac{\partial H}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(k|k-1)} \tag{19}$$

The matrix $\mathbf{P}(k|k-1)$ represents the uncertainty in the estimated parameters, and the matrix $\mathbf{H}(k)$ is the the gradient of $H(\cdot)$ with respect to the parameters evaluated with $\mathbf{P}(k-1|k-1)$. The model uncertainty can be represented as follows,

$$\sigma_y^2 = var(\hat{\mathbf{y}}) = \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) \tag{20}$$

Substituting Eq.(15b), the approximation uncertainty in Eq.(8) can be rewritten as follows,

$$var(\mathbf{y}(N+1)|\mathbf{x}_{N+1}) = \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k) \tag{21}$$

Combining the nonlinear system in Eq.(14) with Eq.(16a) and Eq.(18), we have

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}(k|k-1) \\ \mathbf{y}(k) - H\left(\hat{\boldsymbol{\theta}}(k|k-1)\right) + \mathbf{H}(k)\hat{\boldsymbol{\theta}}(k|k-1) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{I}_{N_h \times N_h} \\ \mathbf{H}(k) \end{bmatrix} \boldsymbol{\theta}(k) + \mathbf{v}(k) \tag{22}$$

where,

$$\mathbf{v}(k) = \begin{bmatrix} -\left(\boldsymbol{\theta}(k) - \hat{\boldsymbol{\theta}}(k|k-1)\right) \\ \mathbf{r}(k) \end{bmatrix} \tag{23}$$

with

$$cov(\mathbf{v}(k)) = \begin{bmatrix} \mathbf{P}(k|k-1) & 0 \\ 0 & \hat{\mathbf{R}}(k) \end{bmatrix} \tag{24}$$

where, $\hat{\mathbf{R}}(k)$ is the estimation of the covariance matrix of innovation residual to be consistent with its theoretical covariance adaptively. This is obtained by averaging the previous residual sequence over the window length N_w,

$$\mathbf{C}_e = \frac{1}{N_w} \sum_{s=k_0}^{k} \left(\mathbf{y}(s) - \mathbf{H}(s)\hat{\boldsymbol{\theta}}(s|s-1)\right)^T \left(\mathbf{y}(s) - \mathbf{H}(s)\hat{\boldsymbol{\theta}}(s|s-1)\right) \tag{25}$$

where $k_0 = k - N_w + 1$. The estimated measurement noise is computed by comparing the theoretical covariance $\mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k)$ with the estimated covariance as follow,

$$\hat{\mathbf{R}}(k+1) = \mathbf{C}_e - \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) \qquad (26)$$

Hence, the approximation uncertainty in Eq.(21) can be rewritten as

$$var(\mathbf{y}(N+1)|\mathbf{x}_{N+1}) = \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \hat{\mathbf{R}}(k) \qquad (27)$$

Remark 3. *In the measurement noise variance, for example, the initialization of the sensor error states in the family of Kalman filter algorithms is the key step to meet the time-varying property. When estimating sensor errors, a low measurement noise variance will result in a precise yet most likely biased estimate. This will result in a long transition to the correct error estimate. Conversely, a larger a priori estimate of measurement noise will result a quicker transition to the correct error estimate but will result in a less precise estimate. By adapting the process matrix in the Kalman filter, both characteristics can be utilized to result in a quick transition to a precise unbiased estimate. Therefore, it is necessary to adapt the measurement noise covariance $\mathbf{R}(k)$ to accommodate for changes in dynamic model and environmental conditions.*

B. The optimal estimate solution of parameters

Traditional Kalman filter works well under Gaussian noises, but it will be deteriorated significantly under non-Gaussian noises, especially when the underlying system is disturbed by impulsive noises, since the Kalman filter captures only the second order statistics of the error signal and is sensitive to the large outliers. The maximum correntropy criterion can be integrated to address this problem, which inherently contains second and higher order moments of the error.

Through Cholesky decomposition, $cov(\mathbf{v}(k))$ can be further described as

$$\begin{aligned} &cov(\mathbf{v}(k)) \\ &= \begin{bmatrix} \mathbf{B}_P(k|k-1)\mathbf{B}_P^T(k|k-1) & 0 \\ 0 & \mathbf{B}_r(k)\mathbf{B}_r^T(k) \end{bmatrix} \\ &= \mathbf{B}(k)\mathbf{B}^T(k) \end{aligned} \qquad (28)$$

where $\mathbf{B}(k)$ can be obtained by Cholesky decomposition of $cov(v(k))$. Multiplying both sides of Eq.(22) by $\mathbf{B}^{-1}(k)$, we have

$$\mathbf{D}(k) = \mathbf{W}(k)\theta(k) + \boldsymbol{\varepsilon}(k) \qquad (29)$$

where,

$$\boldsymbol{\varepsilon}(k) = \mathbf{B}^{-1}(k)\mathbf{v}(k) \qquad (30)$$

$$\begin{aligned} \mathbf{W}(k) &= \mathbf{B}^{-1}(k)\begin{bmatrix} \mathbf{I} \\ \mathbf{H}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{B}_P^{-1}(k|k-1) & 0 \\ 0 & \mathbf{B}_r^{-1}(k) \end{bmatrix}\begin{bmatrix} \mathbf{I} \\ \mathbf{H}(k) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{B}_P^{-1}(k|k-1) \\ \mathbf{B}_r^{-1}(k)\mathbf{H}(k) \end{bmatrix} \end{aligned} \qquad (31)$$

$$\mathbf{D}(k) = \mathbf{B}^{-1}(k)\begin{bmatrix} \hat{\theta}(k|k-1) \\ \mathbf{y}(k) - H\left(\hat{\theta}(k|k-1)\right) + \mathbf{H}(k)\hat{\theta}(k|k-1) \end{bmatrix} \qquad (32)$$

Since $cov(\boldsymbol{\varepsilon}(k)) = cov\left(\mathbf{B}^{-1}(k)\mathbf{v}(k)\right) = \mathbf{B}^{-1}(k)cov(\mathbf{v}(k))(\mathbf{B}^{-1})^T = \mathbf{I}$, the residual error $\boldsymbol{\varepsilon}(k)$ are white.

We propose the following maximum correntropy criterion based cost function

$$J(k) = \frac{1}{L}\sum_{l=1}^{L} G_\sigma\left(d_l(k) - \mathbf{w}_l(k)\theta(k)\right) \qquad (33)$$

where, $d_l(k)$ is the lth element of $\mathbf{D}(k)$, $\mathbf{w}_l(k)$ is the lth row of $\mathbf{W}(k)$. Then the optimal estimate of $\theta(k)$ is

$$\hat{\theta}(k|k) = \arg\max_\theta J(k) = \arg\max_\theta \sum_{l=1}^{L} G_\sigma(\varepsilon_l(k)) \qquad (34)$$

where, $\varepsilon_l(k) = d_l(k) - \mathbf{w}_l(k)\hat{\theta}(k|k)$.

Remark 4. *Different from the well-known mean square error consisting only second order moment, correntropy can be described by a weighted sum of all even order moments, i.e., second order and higher order moments, resulting from the Taylor series expansion of the Gaussian kernel. These higher moments make it possible to be used in robust adaptive filtering in impulsive (heavy-tailed) noise environments.*

The solution of the $\hat{\theta}(k|k)$ can be found by solving

$$\frac{\partial J(k)}{\partial \hat{\theta}(k|k)} = 0 \qquad (35)$$

Then we have

$$\sum_{l=1}^{L} G_\sigma(\varepsilon_l(k))\left(d_l(k) - \mathbf{w}_l(k)\hat{\theta}(k|k)\right)\mathbf{w}_l^T(k) = 0 \qquad (36)$$

It is easy to get

$$\begin{aligned} \hat{\theta}(k|k) &= \left(\sum_{l=1}^{L}\mathbf{w}_l^T(k)G_\sigma(\varepsilon_i(k))\mathbf{w}_l(k)\right)^{-1}\left(\sum_{l=1}^{L}G_\sigma(\varepsilon_l(k))d_l(k)\mathbf{w}_l^T(k)\right) \\ &= \left(\mathbf{W}^T(k)\mathbf{C}(k)\mathbf{W}(k)\right)^{-1}\left(\mathbf{W}^T(k)\mathbf{C}(k)\mathbf{D}(k)\right) \end{aligned} \qquad (37)$$

where,

$$\mathbf{C}(k) = \begin{bmatrix} \mathbf{C}_x(k) & 0 \\ 0 & \mathbf{C}_y(k) \end{bmatrix} \qquad (38)$$

with $\mathbf{C}_x(k) = diag\left(G_\sigma(\varepsilon_1(k)), \ldots, G_\sigma(\varepsilon_{N_h N_y}(k))\right)$, and $\mathbf{C}_y(k) = diag\left(G_\sigma(\varepsilon_{N_h N_y+1}(k)), \ldots, G_\sigma(\varepsilon_{N_h N_y+N_y}(k))\right)$.

Substituting Eq.(31) and Eq.(38), we have

$$\begin{aligned} &\left(\mathbf{W}^T(k)\mathbf{C}(k)\mathbf{W}(k)\right)^{-1} \\ &= \left(\begin{bmatrix} \mathbf{B}_P^{-1}(k|k-1) \\ \mathbf{B}_r^{-1}(k)\mathbf{H}(k) \end{bmatrix}^T\begin{bmatrix} \mathbf{C}_x(k) & 0 \\ 0 & \mathbf{C}_y(k) \end{bmatrix}\begin{bmatrix} \mathbf{B}_P^{-1}(k|k-1) \\ \mathbf{B}_r^{-1}(k)\mathbf{H}(k) \end{bmatrix}\right)^{-1} \\ &= \left((\mathbf{B}_P^{-1}(k|k-1))^T\mathbf{C}_x(k)\mathbf{B}_P^{-1}(k|k-1) \right. \\ &\quad \left. + \mathbf{H}^T(k)(\mathbf{B}_r^{-1})^T(k)\mathbf{C}_y(k)\mathbf{B}_r^{-1}(k)\mathbf{H}(k)\right)^{-1} \end{aligned} \qquad (39)$$

Using the following matrix inversion lemma,

$$(\mathcal{A} + C\mathcal{B}C^T)^{-1} = \mathcal{A}^{-1} - \mathcal{A}^{-1}C(\mathcal{B}^{-1} + C^T\mathcal{A}^{-1}C)^{-1}C^T\mathcal{A}^{-1} \qquad (40)$$

and defining

$$\mathcal{A} = \left(\mathbf{B}_P^{-1}(k|k-1)\right)^T\mathbf{C}_x(k)\mathbf{B}_P^{-1}(k|k-1) \qquad (41a)$$

$$(\mathbf{W}^T(k)\mathbf{C}(k)\mathbf{W}(k))^{-1}$$

$$= \left(\left(\mathbf{B}_P^{-1}(k|k-1)\right)^T \mathbf{C}_x(k)\mathbf{B}_P^{-1}(k|k-1)\right)^{-1} - \left(\left(\mathbf{B}_P^{-1}(k|k-1)\right)^T \mathbf{C}_x(k)\mathbf{B}_P^{-1}(k|k-1)\right)^{-1} \mathbf{H}^T(k)\left(\left(\left(\mathbf{B}_r^{-1}(k)\right)^T \mathbf{C}_y(k)\mathbf{B}_r^{-1}(k)\right)^{-1}\right.$$

$$\left. + \mathbf{H}(k)\left(\left(\mathbf{B}_P^{-1}(k|k-1)\right)^T \mathbf{C}_x(k)\mathbf{B}_P^{-1}(k|k-1)\right)^{-1}\mathbf{H}^T(k)\right)^{-1} \mathbf{H}(k)\left(\left(\mathbf{B}_P^{-1}(k|k-1)\right)^T \mathbf{C}_x(k)\mathbf{B}_P^{-1}(k|k-1)\right)^{-1}$$

$$= \mathbf{B}_P(k|k-1)\mathbf{C}_x^{-1}(k)\left(\mathbf{B}_P(k|k-1)\right)^T - \mathbf{B}_P(k|k-1)\mathbf{C}_x^{-1}(k)\left(\mathbf{B}_P(k|k-1)\right)^T \mathbf{H}^T(k)$$

$$\left(\mathbf{B}_r(k)\mathbf{C}_y^{-1}(k)\mathbf{B}_r^T(k) + \mathbf{H}(k)\mathbf{B}_P(k|k-1)^T \mathbf{C}_x^{-1}(k)\mathbf{B}_P^T(k|k-1)\mathbf{H}^T(k)\right)^{-1} \mathbf{H}(k)\mathbf{B}_P(k|k-1)\mathbf{C}_x^{-1}(k)\mathbf{B}_P^T(k|k-1)$$

$$= \bar{\mathbf{P}}(k|k-1) - \bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) + \mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\right)^{-1} \mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)$$

$$(43)$$

$$\mathcal{B} = \left(\mathbf{B}_r^{-1}\right)^T(k)\mathbf{C}_y(k)\mathbf{B}_r^{-1}(k) \tag{41b}$$

$$\mathcal{C} = \mathbf{H}^T(k) \tag{41c}$$

we can further obtain Eq.(43), where

$$\bar{\mathbf{P}}(k|k-1) = \mathbf{B}_P(k|k-1)\mathbf{C}_x^{-1}(k)(\mathbf{B}_P(k|k-1))^T \tag{44a}$$

$$\bar{\mathbf{R}}(k) = \mathbf{B}_r(k)\mathbf{C}_y^{-1}(k)\mathbf{B}_r^T(k) \tag{44b}$$

Substituting Eq.(31) and Eq.(38), we have

$$\mathbf{W}^T(k)\mathbf{C}(k)\mathbf{D}(k)$$

$$= \begin{bmatrix} \mathbf{B}_P^{-1}(k|k-1) \\ \mathbf{B}_r^{-1}(k)\mathbf{H}(k) \end{bmatrix}^T \begin{bmatrix} \mathbf{C}_x(k) & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_y(k) \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{B}_P^{-1}(k)\hat{\theta}(k|k-1) \\ \mathbf{B}_r^{-1}(k)\left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1)\right) + \mathbf{H}(k)\hat{\theta}(k|k-1)\right) \end{bmatrix}$$

$$= \left(\mathbf{B}_P^{-1}(k|k-1)\right)^T \mathbf{C}_x(k)\mathbf{B}_P^{-1}(k)\hat{\theta}(k|k-1) + \mathbf{H}^T(k)\left(\mathbf{B}_r^{-1}(k)\right)^T$$

$$\mathbf{C}_y(k)\mathbf{B}_r^{-1}(k)\left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1)\right) + \mathbf{H}(k)\hat{\theta}(k|k-1)\right)$$

$$= \left(\bar{\mathbf{P}}(k|k-1)\right)^{-1} \hat{\theta}(k|k-1) + \mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k)\right)^{-1}$$

$$\left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1)\right) + \mathbf{H}(k)\hat{\theta}(k|k-1)\right)$$

$$(45)$$

Therefore, Eq.(37) can be further written as Eq.(46), where

$$\bar{\mathbf{K}}(k) = \bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) + \mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\right)^{-1} \tag{47}$$

Here, the term in Eq.(46) can be rewritten as

$$\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k)\right)^{-1} - \bar{\mathbf{K}}(k)\mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k)\right)^{-1}$$

$$= \bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k)\right)^{-1} - \left(\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) - \bar{\mathbf{K}}(k)\bar{\mathbf{R}}(k)\right)\left(\bar{\mathbf{R}}(k)\right)^{-1}$$

$$= \bar{\mathbf{K}}(k)$$

$$(48)$$

Thus, the posterior state mean in this case is updated as,

$$\hat{\theta}(k|k) = \hat{\theta}(k|k-1) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\hat{\theta}(k|k-1)$$

$$+ \bar{\mathbf{K}}(k)\left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1)\right) + \mathbf{H}(k)\hat{\theta}(k|k-1)\right)$$

$$= \hat{\theta}(k|k-1) + \bar{\mathbf{K}}(k)\left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1)\right)\right)$$

$$= \hat{\theta}(k|k-1) + \bar{\mathbf{K}}(k)\mathbf{e}_k$$

$$(49)$$

Meanwhile, the posterior corresponding covariance matrix is updated as Eq.(50), where $\mathbf{S}(k) = \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k)$ is the theoretical covariance.

Remark 5. *Since the theoretical covariance $\mathbf{S}(k)$ is also utilized in the null hypothesis $\mathcal{H}_0$ regarded as an estimate*

interval of the uncertainty in the expected output for the given input observation, it is not necessary to repeat computing in the parameter updating process saving the computation resources.

To minimize $tr\left(\mathbf{P}(k|k)\right)$, we have

$$\frac{\mathrm{d}tr\left(\mathbf{P}(k|k)\right)}{\mathrm{d}\left(\bar{\mathbf{K}}(k)\right)} = -2\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \bar{\mathbf{K}}(k)\mathbf{S}(k) = 0 \tag{51}$$

Thus, it can be obtained that

$$\bar{\mathbf{K}}(k)\mathbf{S}(k)\bar{\mathbf{K}}^T(k) = \mathbf{P}(k|k-1)\mathbf{H}^T(k)\bar{\mathbf{K}}^T(k) \tag{52}$$

Substituting Eq.(52) to Eq.(50), the posterior corresponding covariance matrix is rewritten as

$$\mathbf{P}(k|k) = \mathbf{P}(k|k-1) - \mathbf{P}(k|k-1)\mathbf{H}^T(k)\bar{\mathbf{K}}^T(k)$$

$$- \bar{\mathbf{K}}(k)\mathbf{H}(k)\mathbf{P}(k|k-1) + \bar{\mathbf{K}}(k)\mathbf{S}(k)\bar{\mathbf{K}}^T(k)$$

$$= \mathbf{P}(k|k-1) - \mathbf{P}(k|k-1)\mathbf{H}^T(k)\bar{\mathbf{K}}^T(k)$$

$$- \bar{\mathbf{K}}(k)\mathbf{H}(k)\mathbf{P}(k|k-1) + \mathbf{P}(k|k-1)\mathbf{H}^T(k)\bar{\mathbf{K}}^T(k)$$

$$= \mathbf{P}(k|k-1) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\mathbf{P}(k|k-1)$$

$$= (\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k))\mathbf{P}(k|k-1)$$

$$(53)$$

The AMCEKF algorithm is summarized in Algorithm 1.

Remark 6. *The while loop in the parameter updating process represents a fixed-point iterative algorithm. It mainly is used to obtain the optimal $\hat{\theta}(k|k)$ at time k since the parameter $\theta(k|k)$ at time k satisfies the solution expression $\hat{\theta}(k|k) = f(\hat{\theta}(k|k))$ according to Eq.(37). In the iterative procedure, the optimal $\hat{\theta}(k|k)$ would be achieved by updating successively its value based on Eqs.(54)~(60) and the latest estimate $\hat{\theta}(k|k-1)$. According to the fixed-point solution principle, we can see that the parameter updating process only uses the sample at time k without using all past data samples. This indicates that parameter updating process is a sequential learning process. The for loop in Algorithm 1 gives the online sequential learning process of SEFIS. Also, we can observe that the while loop is embedded inside the for loop, that is to say, for each data arriving, the while loop would be invoked and aborted as an alternative efficient way to solve the optimal solution $\hat{\theta}(k|k)$. In this case, the proposed algorithm is an online model.*

V. LEARNING PROCESS OF SEFIS

The learning process of SEFIS is presented based on the above statistical criterion and parameters updating algorithm, which incorporates the advantages of using AMCEKF for estimation and evolving criterion that detects model insufficiency.

$$\hat{\theta}(k|k)$$

$$= \left(\bar{\mathbf{P}}(k|k-1) - \bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) + \mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) \right)^{-1} \mathbf{H}(k)\bar{\mathbf{P}}(k|k-1) \right)$$
$$\left(\left(\bar{\mathbf{P}}(k|k-1) \right)^{-1} \hat{\theta}(k|k-1) + \mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) \right)^{-1} \left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1) \right) + \mathbf{H}(k)\hat{\theta}(k|k-1) \right) \right)$$
$$= \left(\bar{\mathbf{P}}(k|k-1) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\bar{\mathbf{P}}(k|k-1) \right) \left(\left(\bar{\mathbf{P}}(k|k-1) \right)^{-1} \hat{\theta}(k|k-1) + \mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) \right)^{-1} \left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1) \right) + \mathbf{H}(k)\hat{\theta}(k|k-1) \right) \right)$$
$$= \hat{\theta}(k|k-1) + \bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) \right)^{-1} y(k) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\hat{\theta}(k|k-1) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{R}(k) \right)^{-1}$$
$$\left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1) \right) + \mathbf{H}(k)\hat{\theta}(k|k-1) \right)$$
$$= \hat{\theta}(k|k-1) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\hat{\theta}(k|k-1) + \left(\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) \right)^{-1} - \bar{\mathbf{K}}(k)\mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) \right)^{-1} \right)$$
$$\left(\mathbf{y}_k - H\left(\hat{\theta}(k|k-1) \right) + \mathbf{H}(k)\hat{\theta}(k|k-1) \right) \tag{46}$$

$$\mathbf{P}(k|k) = cov\left(\theta(k) - \hat{\theta}(k|k) \right)$$
$$= cov\left(\theta(k) - \left(\hat{\theta}(k|k-1) + \bar{\mathbf{K}}(k)\left(\theta(k|k) - H\left(\hat{\theta}(k|k-1) \right) \right) \right) \right)$$
$$= cov\left(\theta(k) - \left(\hat{\theta}(k|k-1) + \bar{\mathbf{K}}(k)\left(\mathbf{H}(k)\theta(k) + \mathbf{r}(k) - \mathbf{H}(k)\hat{\theta}(k|k-1) \right) \right) \right)$$
$$= cov\left(\left(\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k) \right)\left(\theta(k) - \hat{\theta}(k|k-1) \right) - \bar{\mathbf{K}}(k)\mathbf{r}(k) \right)$$
$$= cov\left(\left(\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k) \right)\left(\theta(k) - \hat{\theta}(k|k-1) \right) \right) + cov\left(\bar{\mathbf{K}}(k)\mathbf{r}(k) \right) \tag{50}$$
$$= \left(\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k) \right)cov\left(\theta(k) - \hat{\theta}(k|k-1) \right)\left(\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k) \right)^T + \bar{\mathbf{K}}(k)cov\left(\mathbf{r}(k) \right)\bar{\mathbf{K}}^T(k)$$
$$= \left(\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k) \right)\mathbf{P}(k|k-1)\left(\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k) \right)^T + \bar{\mathbf{K}}(k)\mathbf{R}(k)\bar{\mathbf{K}}^T(k)$$
$$= \mathbf{P}(k|k-1) - \mathbf{P}(k|k-1)\mathbf{H}^T(k)\bar{\mathbf{K}}^T(k) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\mathbf{P}(k|k-1) + \bar{\mathbf{K}}(k)\mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k)\bar{\mathbf{K}}^T(k) + \bar{\mathbf{K}}(k)\mathbf{R}(k)\bar{\mathbf{K}}^T(k)$$
$$= \mathbf{P}(k|k-1) - \mathbf{P}(k|k-1)\mathbf{H}^T(k)\bar{\mathbf{K}}^T(k) - \bar{\mathbf{K}}(k)\mathbf{H}(k)\mathbf{P}(k|k-1) + \bar{\mathbf{K}}(k)\mathbf{S}(k)\mathbf{K}^T(k)$$

A. Recruiting of Fuzzy Rules

The growth criterion developed above is independent of the model structure, even though it is demonstrated on the evolving fuzzy interference system, and hence is applicable to any type of model. In this paper, the statistical approximation sufficiency is used as the criterion to determine whether a rule needs to be generated. Then the selected type of basic rule to be recruited is the Gaussian membership functions. When the first input-output data $\mathbf{x}_1, \mathbf{y}_1$ sample comes in, the first new rule is generated with the mean and center of the new fuzzy set and its consequent parameters assigned by

$$\mathbf{a}_{N_h+1}^k = \mathbf{y}_1, \boldsymbol{\mu}_1^k = \mathbf{x}_1, \sigma_{N_h+1}^k = \kappa\|\mathbf{x}_1\| \tag{65}$$

When the model sufficiency null hypothesis is rejected for the kth observation in the region of rejection which is described as follows

$$\|\mathbf{e}_k\| > z_d\,\|\mathbf{S}(k)\| \tag{66}$$

the number of the fuzzy rules is set to $N_h + 1$. The region of rejection for null hypothesis $\mathcal{H}_0$ signifies that the statistical model sufficiency cannot be satisfied. For the newly recruited rules, the parameters are assigned as follows

$$\mathbf{a}_{N_h+1}^k = \mathbf{e}_k, \boldsymbol{\mu}_{N_h+1}^k = \mathbf{x}_k, \sigma_{N_h+1}^k = \kappa\|\mathbf{x}_k - \boldsymbol{\mu}_{nr}^k\| \tag{67}$$

where, κ is an overlap factor that determines the overlap of fuzzy rules in the input space, $\boldsymbol{\mu}_{nr}^k$ is the center of the fuzzy rule nearest to $\mathbf{x}_k$. The parameter vector simply tags the new parameters to the existing ones,

$$\theta(k) = \left[\mathbf{a}_1^{kT}, \ldots, \mathbf{a}_{N_h}^{kT}, \mathbf{a}_{N_h+1}^{kT} \right]^T \tag{68}$$

and the parameter error covariance matrix becomes

$$\bar{\mathbf{P}}(k|k) = \begin{pmatrix} \bar{\mathbf{P}}(k|k-1) & \mathbf{0} \\ \mathbf{0} & p_0\mathbf{I}_{N_y \times N_y} \end{pmatrix} \tag{69}$$

where p_0 is an estimate of the uncertainty in the initial values assigned to the parameters which is set 1.0 for all the examples.

Eq.(65) shows that the antecedent parameters are evolved from scratch in an evolvable way. For any new sample, a new rule is recruited when the criteria of adding rule are satisfied as shown in Eq.(66) and then its antecedent parameters are determined with the newly loaded sample based on Eq.(67). In this way, the antecedent parameters are permanently updated whenever a new rule is generated. In Eq.(66), the statistical model sufficiency criterion ensures that the new rule is added at the time of architecture deficiency, which could be effectively avoided to passively increase scale by the influence of the outliers.

B. Updating of Fuzzy Rules

In the SEFIS, the antecedent parameters of the rules are determined according to Eq.(67) in an evolvable way. When there is no addition of rules, the consequent parameters are updated. The parameter vector existing in all the fuzzy rules

---

Algorithm 1 Adaptive Maximum Correntropy Kalman Filter

Set the initial estimate $\mathbf{P}(0|0)$, $\hat{\theta}(0|0)$, $\hat{\mathbf{R}}(1)$.

for Each observation $(\mathbf{x}_k, \mathbf{y}_k)$ **do**

    Compute the prior state mean $\hat{\theta}(k|k-1)$ via Eq.(16a) **and** corresponding covariance matrix $\mathbf{P}(k|k-1)$ via Eq.(16b)

    Use Cholesky decomposition to obtain $\mathbf{B}_P(k|k-1)$ and $\mathbf{B}_r(k)$

    Given $t = 1$, and $\hat{\theta}(k|k)_t = \hat{\theta}(k|k-1)$, where $\hat{\theta}(k|k)_t$ denotes the estimated state at fixed-point iteration t

    while $\dfrac{\|\hat{\theta}(k|k)_t - \hat{\theta}(k|k)_{t-1}\|}{\|\hat{\theta}(k|k)_{t-1}\|} \le \nu$ **do**

$$\hat{\theta}(k|k)_t = \hat{\theta}(k|k-1) + \bar{\mathbf{K}}(k)\mathbf{e}(k) \tag{54}$$

        with

$$\varepsilon_l(k) = d_l(k) - \mathbf{w}_l(k)\hat{\theta}(k|k)_{t-1} \tag{55}$$

$$\mathbf{C}_x(k) = diag\left(G_\sigma\left(\varepsilon_1(k)\right),\ldots,G_\sigma\left(\varepsilon_{N_h N_y}(k)\right)\right) \tag{56}$$

$$\mathbf{C}_y(k) = diag\left(G_\sigma\left(\varepsilon_{N_h N_y+1}(k)\right),\ldots,G_\sigma\left(\varepsilon_{N_h N_y+N_y}(k)\right)\right) \tag{57}$$

$$\bar{\mathbf{P}}(k|k-1) = \mathbf{B}_P(k|k-1)\mathbf{C}_x^{-1}(k)(\mathbf{B}_P(k|k-1))^T \tag{58}$$

$$\bar{\mathbf{R}}(k) = \mathbf{B}_r(k)\mathbf{C}_y^{-1}(k)\mathbf{B}_r^T(k) \tag{59}$$

$$\bar{\mathbf{K}}(k) = \bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\left(\bar{\mathbf{R}}(k) + \mathbf{H}(k)\bar{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)\right)^{-1} \tag{60}$$

$$t \to t + 1$$

    end while

$$\mathbf{P}(k|k) = (\mathbf{I} - \bar{\mathbf{K}}(k)\mathbf{H}(k))\mathbf{P}(k|k-1) \tag{61}$$

$$\mathbf{S}(k) = \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \hat{\mathbf{R}}(k) \tag{62}$$

$$\mathbf{C}_e = \frac{1}{N_w}\sum_{s=k_0}^{k}\left(\mathbf{y}(s)-\mathbf{H}(s)\hat{\theta}(s|s-1)\right)^T\left(\mathbf{y}(s)-\mathbf{H}(s)\hat{\theta}(s|s-1)\right) \tag{63}$$

$$\hat{\mathbf{R}}(k+1) = \mathbf{C}_e - \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) \tag{64}$$

end for

Fig. 1. SEFIS algorithm

is given by $\theta(k) = \left[\mathbf{a}_1^{kT},\ldots,\mathbf{a}_n^{kT},\ldots,\mathbf{a}_{N_h}^{kT}\right]^T$, where $\mathbf{a}_n^{kT}$ is the parameter vector and its gradient is derived as follows,

$$\dot{\mathbf{a}}_n^k = \frac{\partial\hat{\mathbf{y}}_k}{\partial\mathbf{a}_n^k} = \frac{R_n(\mathbf{x}_k)}{\sum_{n=1}^{N_h}R_n(\mathbf{x}_k)}\mathbf{I}_{N_y\times N_y} \tag{70}$$

After obtaining the gradient vector of the parameters, that is $\mathbf{H}(k) = [\dot{\mathbf{a}}_1^k,\ldots,\dot{\mathbf{a}}_n^k\ldots,\dot{\mathbf{a}}_{Nh}^k]$, AMCEKF is utilized to update the parameters $\theta(k)$.

In AMCEKF, the system output is ensured to be close to the desired response through maximizing the correntropy of the synthetic of the process error and measurement error in Kalman filter with an adaption process. Also, the AMCEKF can be applied in any noise environments which is insensitive to the peak in the noise, and effectively handle the bulk of residuals around the origin. Thus, the AMCEKF being a Kalman filter-like recursion algorithm outperform the original extended Kalman filter significantly, especially when the underlying system is disturbed by some impulsive non-Gaussian noises.

C. Pruning the fuzzy rule

If the null hypothesis for the statistical inference of system sufficiency is satisfied as

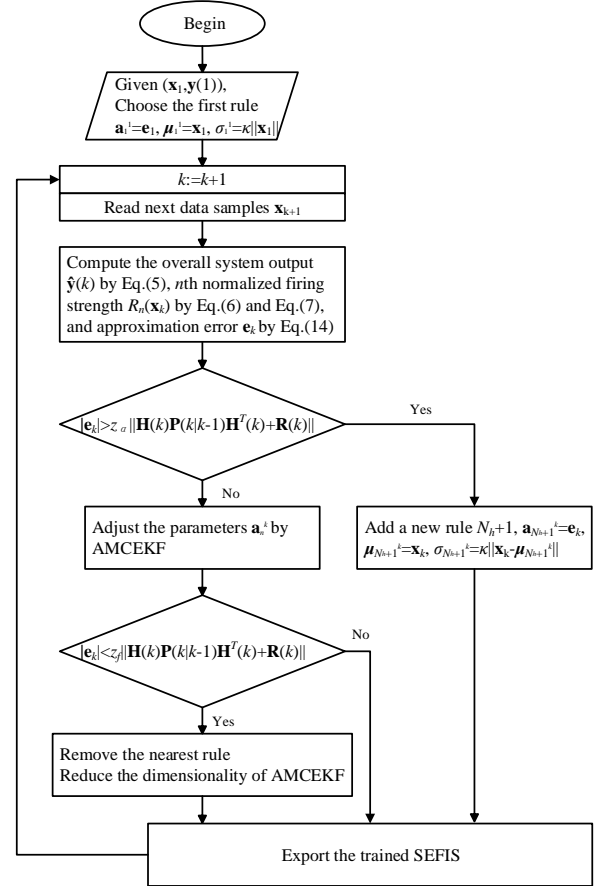$$\|\mathbf{e}_k\| < z_f\|\mathbf{S}(k)\| \quad \text{for } f\% \text{ of data with } f < d \tag{71}$$

where, f is usually designed small enough. That is to say that, the system structure has the appropriate quantities of rules or nodes, especially when the approximation error converges to a minor constructed interval relative to the approximation uncertainty. The pruning strategy searches for the minimum redundancy as well as avoids deleting the components that most dominate the interpretability by calculating membership functions. The nearest rule $\bar{R}_{nr}$ presents the inactive rule under the current data stream and further would be deleted without losing any information and accuracy of the subsequent trained models when the model sufficiency is redundant.

The SEFIS algorithm is summarized in Fig. 1.

D. Computational complexity

The SEFIS algorithm mainly involves recruiting and pruning of rules and AMCEKF for parameter updating. Besides, the recruiting and pruning of rules mainly depends on the statistical sufficiency criterion, that can be also obtained by AMCEKF. Thus, the computational complexity of the proposed SEFIS mainly relies on the AMCEKF calculation procedure. We analyze its computational complexity in terms of the floating point operations. The computational complexities for the equations used in AMCEKF are given in Table I. Assuming that the average fixed-point iteration number is T that is relatively small in general, based on Table I the upper limit of the computational complexity is obtained as,

TABLE I
Computational complexities of some equations

Equation	Addition/subtraction and multiplication	Division, matrix inversion, Cholesky decomposition and exponentiation
Eq.(16a)	$2\mathcal{P}^2 - \mathcal{P}$	0
Eq.(16b)	$4\mathcal{P}^3 - \mathcal{P}^2$	0
Eq.(54)	$2\mathcal{P}Q$	0
Eq.(55)	$2Q$	0
Eq.(56)	$2\mathcal{P}^2$	$\mathcal{P}$
Eq.(57)	$2Q^2$	Q
Eq.(58)	$2\mathcal{P}^2 + \mathcal{P}$	$O(\mathcal{P}^3)$
Eq.(59)	$2Q^2 + Q$	$O(Q^3)$
Eq.(60)	$4\mathcal{P}^2 Q + 4Q^2\mathcal{P} - 3\mathcal{P}Q$	$O(Q^3)$
Eq.(61)	$2\mathcal{P}^2 Q + 2\mathcal{P}^3 - \mathcal{P}^2$	0
Eq.(62)	$2\mathcal{P}^2 Q + 2Q^2\mathcal{P} - \mathcal{P}Q$	0
Eq.(63)	$(2\mathcal{P}Q + Q^2)N_w$	0
Eq.(64)	$2\mathcal{P}^2 Q + 2Q^2\mathcal{P} - \mathcal{P}Q$	0

For simplicity, $N_h N_y = \mathcal{P}$, $N_y = Q$.

$$S_{SEFIS} = 6\mathcal{P}^3 + (6+4T)\mathcal{P}^2 Q + (4+4T)Q^2\mathcal{P} + (2T-1)\mathcal{P}$$
$$+ (2N_w - 2 - T)\mathcal{P}Q + (N_w + 4T)Q^2 + 4TQ \quad (72)$$
$$+ 4T\mathcal{P}^2 + 2TO(Q^3) + TO(\mathcal{P}^3)$$

Eq.(72) indicates that the computational complexity is obviously with the number of rules and it can be efficiently reduced by restricting the size of rules under the statistical sufficiency criterion.

VI. Performance Evaluation of SEFIS

In this section, the performance of the proposed SEFIS is evaluated by a nonlinear system identification problems, some regression problem, the adaptive noise cancellation, and the Typhoon tracking prediction. The first two examples have been added some impulse noises deliberately to verify the rejection property of non-Gaussian noises. The latter two are endowed with the real non-Gaussian noises, which require new targeted solution to improve the performance under the complicated noise conditions. For purposes of comparison, other fuzzy neural system algorithms, such as SAFIS[17], ESAFIS [18], OS-Fuzzy-ELM[25], eTS[15], and Simpl_eTS[16], CEFNS[20], RMCEFS[21], et.al., are utilized to further demonstrate the performance superiority.

A. Nonlinear System Identification

The nonlinear system [26, 27] to be identified is described as

$$y(t+1) = \frac{y(t)y(t-1)[y(t) + 2.5]}{1 + y^2(t) + y^2(t-1)} + u(t) \quad (73)$$

If a series-parallel identification model is used for identifying the system, the model can be described as follows,

$$\hat{y}(t+1) = f(y(t), y(t-1), u(t)) \quad (74)$$

where f is the function implemented to the SEFIS with three inputs and one output.

The input and output data have been collected in such a way that the input is given as $u(t) = \sin(2\pi t/25)$. For the purposes of training and testing, 500 samples are produced. Also to further assess the non-Gaussian noise resistance performance

of the SEFIS, the impulse noise with amplitude 2 at an interval of ten samples are added in the input data. Another common type of non-Gaussian white noise called non-Gaussian white noise is considered in the nonlinear system identification. To verify the statistical performance of the testing results, we uses the bootstrapping for the dataset of nonlinear identification system as an example. Bootstrapping is a test or metric that relies on random sampling with replacement. We randomly choose 300 samples each time for testing, and the other are used to train the system. The mean results are achieved by repeating 10 times and shown in Table II.

The performance comparison between different algorithms under the noise-free, non-Gaussian white noise, and impulse noise conditions are given in Table II. From Table II, one can find that the proposed SEFIS obtains better testing accuracy with lesser training time and number of rules in the impulse noise condition. In noise-free condition, the ESAFIS and OS-Fuzzy-ELM with less rules perform better accuracy than SEFIS, however, they cost more training time in the learning process than SEFIS. This further verifies the robust and noise rejection abilities of the proposed SEFIS. Besides, the table reveals that the proposed SEFIS achieves better performance than the CEFNS and RMCEFS based on the proposed rule revolution and parameter updating methods in rejecting outliers.

The training process of the nonlinear system identification using the first 300 samples as an example to indicate the statistical significance is shown in Figs.2. From Fig.2(b), one can see that the identification errors under non-Gaussian noise would be very volatile and high, and the continuous real-time observation of statistical approximation uncertainty is utilized to measure the model sufficiency. If the conventional error criterion with a threshold is used to recruit the new rules, the sizes of the rules would be either too large with a relatively small threshold, or zero growth with a large threshold. Both cases are not desirable. The statistical model sufficiency gives the perfect solution. At the beginning of the learning process, the error lies beyond the constructed interval of the proposed statistical approximation uncertainty and some rules are added consequentially. And then the statistical approximation uncertainty becomes relative stability and the current error falls inside the interval with a certain level of confidence. At about 170th time step, the system appears to obtain appropriate quantities of rules which manifests smaller approximation uncertainty shown in Fig.2(c). Thus, when the current error lies in a minor constructed interval relative to the approximation uncertainty, the pruning strategy would be triggered shown in Fig.2(d).

B. Regression

In this example, five real-world regression data sets [28] are further considered to validate the performance of SEFIS. For all the data sets, the input and output attributes are normalized into the range [0 1]. The testing error as well as the training time and number of rules for different algorithms are given in Table III. The table shows that the proposed SEFIS achieves better generalization than SAFIS, ESAFIS, OS-Fuzzy-ELM,

TABLE II
PERFORMANCE COMPARISON FOR NONLINEAR SYSTEM IDENTIFICATION

Algorithms	Without Noise			With non-Gaussian White Noise			With Impulse Noise		
	Testing RMSE	Training Time	Rules	Testing RMSE	Training Time	Rules	Testing RMSE	Training Time	Rules
SAFIS	0.2097	0.3438	6	0.1887	0.7344	10	0.2973	1.2500	13
ESAFIS	0.0686	0.5000	8	0.1780	0.4375	10	0.3592	5.1094	12
OS-Fuzzy-ELM	**0.0374**	0.1406	**3**	0.1967	1.0469	20	0.4389	0.9688	15
eTS	0.2124	0.2188	5	0.1971	0.2969	9	0.2936	**0.1875**	8
Simpl_eTS	0.1972	0.2969	6	0.1972	0.3281	9	0.2906	0.3428	35
CEFNS	0.1731	0.5156	12	0.1731	0.5939	15	0.2026	0.4844	11
RMCEFS	0.0969	0.8594	10	0.1652	1.5313	19	0.1945	0.2928	12
SEFIS	**0.0952**	**0.1204**	**8**	**0.1399**	**0.2415**	**8**	**0.1487**	**0.1979**	**6**

(a) Identification performance

(b) Identification errors

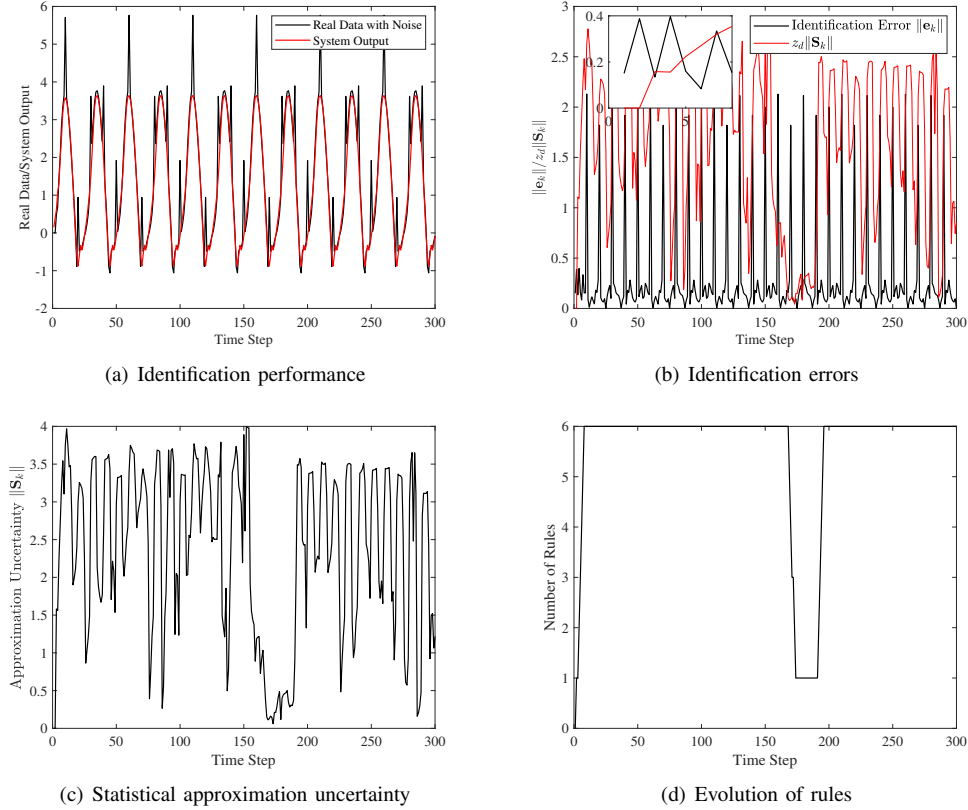(c) Statistical approximation uncertainty

(d) Evolution of rules

Fig. 2. Identification results

eTS, Simpl_eTS and CEFNS with the least testing error and shortest training time even in the noise-free case. Nevertheless, it achieves much close learning accuracy with RMCEFS in less time. Under the impulse circumstances, SEFIS obviously spends less time and obtains much better testing accuracies than all those algorithms with its outlier resistance capability. However, SEFIS requires more rules due to the more stable recruiting and pruning criterion based on the statistical concept with a slower-growing and slower-reducing way. Although the more rules result in the increase of the nominal computation burden, the statistical criterion in minimalism release the cost greatly which can effectively reduce the final training time.

C. Adaptive Noise Cancellation

Consider an adaptive noise canceler shown in Fig.3, where the signal source is $s(k) = \sin(0.06k)\cos(0.01k)$ and the noise signal $n(k)$ is generated by a white noise. After passing through
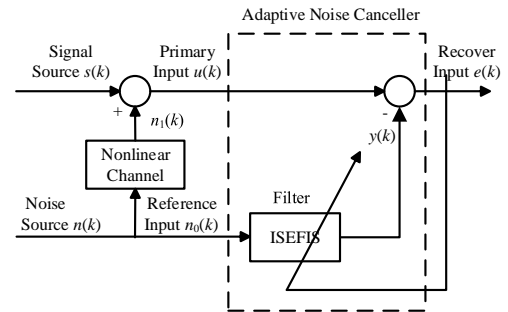

Fig. 3. Structure of adaptive noise cancellation

a nonlinear channel with a nonlinear function, a corrupting noise $n_1(k)$ is created. The relationship between $n(k)$ and $n_1(k)$

TABLE III
RESULT COMPARISON FOR REGRESSION BENCHMARK PROBLEMS

database	Algorithms	Without Noise			With Impulse Noise		
		Testing RMSE	Training Time	Rules	Testing RMSE	Training Time	Rules
Autos	SAFIS	0.1184	0.4524	5	0.1377	0.3120	9
	ESAFIS	0.0604	0.2184	3	0.1290	0.3276	5
	OS-Fuzzy-ELM	0.0595	0.0296	2	0.1612	0.0866	5
	eTS	0.0535	0.2184	3	0.1449	1.1856	19
	Simpl_eTS	0.0689	0.5772	10	0.1208	0.1716	6
	CEFNS	0.0666	0.0156	2	0.0976	0.0312	5
	RMCEFS	0.0409	0.0156	2	0.0654	0.0312	4
	SEFIS	0.0622	0.0156	5	0.0661	0.0156	8
Triazines	SAFIS	0.0581	5.2260	9	0.1006	15.1976	17
	ESAFIS	0.0331	24.4438	19	0.0907	12.4824	13
	OS-Fuzzy-ELM	0.0100	2.4633	6	0.2338	8.6846	8
	eTS	0.0179	3.6816	9	0.1121	14.5237	17
	Simpl_eTS	0.0197	3.5220	9	0.1125	41.8083	25
	CEFNS	0.0452	0.1560	6	0.0647	0.0313	4
	RMCEFS	0.0055	0.1560	4	0.0120	0.1716	4
	SEFIS	0.0068	0.0313	4	0.0102	0.0313	5
Auto MPG	SAFIS	0.0979	0.2188	2	0.1662	0.3791	5
	ESAFIS	0.0679	0.2500	2	0.1335	0.4219	2
	OS-Fuzzy-ELM	0.2037	0.0469	5	0.2346	0.0625	5
	eTS	0.1735	0.6250	5	0.1786	0.6406	5
	Simpl_eTS	0.0993	0.2969	4	0.1602	0.1719	4
	CEFNS	0.0750	0.0781	2	0.0845	0.1710	2
	RMCEFS	0.0672	0.2344	2	0.0763	0.0781	3
	SEFIS	0.0678	0.0313	5	0.0691	0.0313	5
Real Estate Valuation	SAFIS	0.1015	0.4219	13	0.1090	0.4063	12
	ESAFIS	0.1052	0.9219	20	0.1155	0.4642	37
	OS-Fuzzy-ELM	0.0712	1.1875	10	0.1247	0.2072	12
	eTS	0.0886	5.2656	14	0.1016	5.4688	13
	Simpl_eTS	0.0910	3.7031	10	0.1139	3.2188	10
	CEFNS	0.1006	0.5496	15	0.0970	0.6719	15
	RMCEFS	0.0855	2.1875	14	0.0899	8.0938	20
	SEFIS	0.0836	0.1719	9	0.0826	0.3750	8
Yacht Hydrodynamics	SAFIS	0.2890	0.2188	13	0.2950	0.6406	3
	ESAFIS	0.1784	2.1406	12	0.2396	4.6250	18
	OS-Fuzzy-ELM	0.1459	0.6919	8	0.1892	0.7500	10
	eTS	0.1419	0.4844	10	0.1731	0.4063	12
	Simpl_eTS	0.1423	0.3750	12	0.1734	0.4844	11
	CEFNS	0.1732	1.0938	15	0.1686	0.3906	7
	RMCEFS	0.1476	1.6406	19	0.1261	12.9063	26
	SEFIS	0.0986	0.4531	10	0.0990	0.2656	10
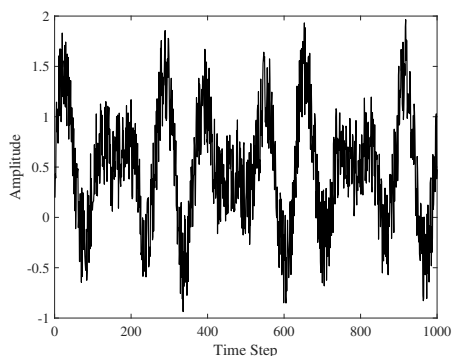
(a) Corrupted signal
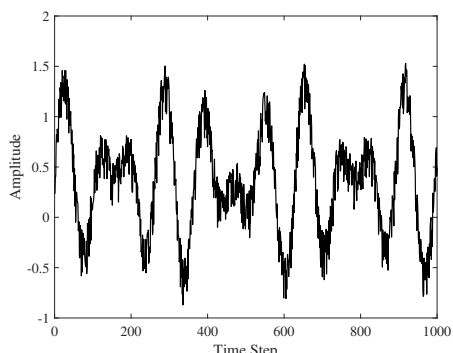
(b) Recovered output

Fig. 4. Simulation results by SEFIS

is represented as $n_1(k) = 0.6n^3(k)$. The reference signal is the input of the SEFIS, that is $n_0(k) = n(k)$. When the output of the SEFIS equals to $n_1(k)$, the noise is canceled. Then, the output of the adaptive noise canceler is the recovered source signal. The number of the training data is 1000. Different from the existing results which achieved from 100 training epochs shown in [29], after only one training epoch, the simulation result is obtained. The compared models are considered to have the same initial values. After 10 independent runs, the average RMSE by the proposed model has a value of 0.0082, which achieves the best filtering performance compared with the five models reported in [29] for this adaptive noise canceler system. To further assess the non-Gaussian noise resistance performance of the SEFIS, the impulse noise with amplitude 1 at an interval of ten training data are added in the input data. The recovered signals after the training by using the SEFIS are shown in Fig.4. Table IV gives the RMSE for the different algorithms under non-Gaussian noise. From the table, SEFIS obtains better filter performance than the other fuzzy algorithms. The results indicate that the performance of the proposed filters is quite well, and the development of nonlinear filters for impulse noise condition is necessary.

D. Typhoon Path Prediction

Due to the dynamical and thermodynamical asymmetric structure, and the environmental systems of different scales, typhoon track prediction is a challenge problem in typhoon research. Accuracy is one of the significant criteria when talking about the prediction of typhoon disasters, as well as the speed and flexibility. Forecasts should be done quickly and forecast tools should be able to react immediately on sudden track disturbances, which is resulted by, for instance, subtropical anticyclone and multi-time scale summer monsoon circulation. These disturbances are different from the noises of underlying surfaces and secondary low pressure, and may lead to the transient abrupt path. In this work, track of typhoons are predicted using the proposed SEFIS with the past track and the other meteorological conditions as inputs taking advantages of the evolvable structure and the rejection property of a broad spectrum of noise environment. Time series of dataset of typhoons provided by National Tibetan Plateau Data Center [30], which occurred in the West-Pacific in the past are used to train the SEFIS automatically. In such a scenario, evaluating the algorithm by dividing the data into training and testing sets is unsuitable. Here, we need to develop an online tracker that the training SEFIS is employed to produce a full life cycle track of a typhoon. The predicted tracking of a typhoon favorably identifies the future location of the typhoon center. Errors between predicted and real typhoon centers are measured quantitatively.

In this study, we predict the typhoon track based on the previous three data, i.e.,

$$Input\begin{cases} Lat(t-2) & Lat(t-1) & Lon(t) \\ Lon(t-2) & Lon(t-1) & Lon(t) \\ Vel(t-2) & Vel(t-1) & Vel(t) \\ Pre(t-2) & Pre(t-1) & Pre(t) \end{cases} Output\begin{cases} Lat(t+1) \\ Lon(t+1) \end{cases} \quad (75)$$

TABLE IV
Comparison of RMSE for different algorithms under non-Gaussian noise

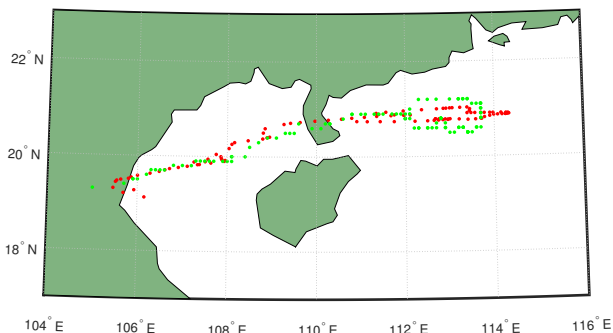| Algorithms	SAFIS	ESAFIS	OS-Fuzzy-ELM	eTs	Simpl_eTS	CEFNS	RMCEFS	SEFIS
RMSE	0.0639	0.0598	0.0504	0.1297	0.1296	0.0791	0.0587	**0.0503**

Fig. 5. The track prediction of Typhoon Bebinca (No. 1816, 12th Aug 2018-17th Aug 2018), red point indicating actual path and green point indicating prediction path

TABLE V
Comparison of RMSE time for different algorithms

Algorithms	Latitude	Longitude
SAFIS	5.2057	1.1531
ESAFIS	2.1606	1.3027
OS-Fuzzy-ELM	1.7578	1.0139
eTS	1.6792	0.8063
Simpl_eTS	1.0358	1.8816
CEFNS	5.6299	2.6846
RMCEFS	0.6745	0.9027
SEFIS	**0.1822**	**0.6479**

A total of 113 data points every hour are collected for this study. The prediction performance of SEFIS for this problem is shown in Fig. 5. This figure show the actual path and the prediction path by SEFIS. From the figure, it could be observed that SEFIS is able to predict the path and sudden changes in complex environmental system, satisfactorily. Even during the period of typhoon genesis ($112°E$-$114°E$), SEFIS is able to give an acceptable tracking performance. Next, we quantitatively evaluate the performance of SEFIS. Table V gives the perdition RMSE for the different algorithms.
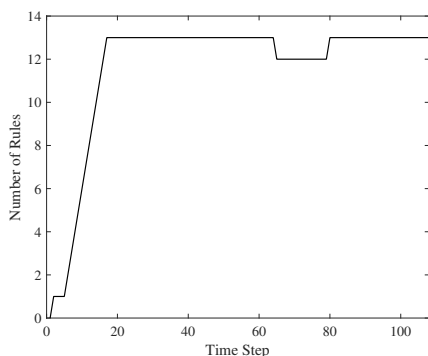

Fig. 6. Evolution of rules during the prediction period

From the table, it could be observed that SEFIS attains better tracking result than the other fuzzy algorithms. The evolution of rules is illustrate in Fig. 6, which begins with an empty fuzzy rule base, gradually increases or decrease the number to capture the complex dynamics of typhoon, and reaches a steady state finally. Also, SEFIS has a smooth rule evolution process despite the prediction is a process of intense uncertainties and sudden changes.

E. State Prediction for the Turbofan Engine

To further demonstrate the availability and performance of the proposed system, the state prediction for the turbofan engine rooted in the industrial application problem of PHM system is tested and verified. In PHM applications, predictor generally aims at computing long-term state evolution from continuous observations in order to detect the fault mode and to finally estimate the remaining useful life of the system.

Turbofan engine dataset gained from NASA prognostic data repository are taken into consideration [31]. The dataset train-FD001.txt contain 21 sensor measurements. The dataset are generated by considering one failure mode of the engines. The dataset are normalized in the range $[0, 1]$. By denoting state **x**, we design the following prediction model,

$$\hat{\mathbf{x}}(k + 1) = f(\mathbf{x}(k), \mathbf{x}(k - 1), \mathbf{x}(k - 2)) \tag{76}$$

where $\mathbf{x}(k) = [x_1(k), \ldots, x_{21}(k)]$ are the states of the 21 sensors.

In this problem, the sensor data are contaminated with noises. The main sources of noises are from manufacturing and assembly variations, process noises, and measurement noises. Fig.7(a) gives the values from sensors 2,8,11,13,15 chosen arbitrarily to show the data characteristics. It can be seen that degradation trajectories shown by sensor states are masked by a mixture of random noises. Research shows that mixture noise models are more difficult to characterize even if they consist of simple individual components. Moreover, the random measurement noise components are mingled with all output channels. This multistage noise contamination results in complex noise characteristics and poses a challenge to carry out appropriate denoising operations. The proposed SEFIS is expected to reflect the trend of engine states hiding in the noise. The continuous prediction for multidimensional future signals is shown in Fig.7(b). From Fig.7(b), it can be found that although the engine continues running with heavy degradation and multistage noise contamination, the inherent noise rejection capacity of the proposed algorithm depicts a relatively high level of satisfactory. Considering the noise interference that masks the degradation trend, the locally weighted scatter plot smoothing [32] as filter is used to denoise and restore the original state trajectories shown in Fig.7(c). Obviously, the state prediction by SEFIS can reflect the long-term state evolution and suppress noise interference effectively.
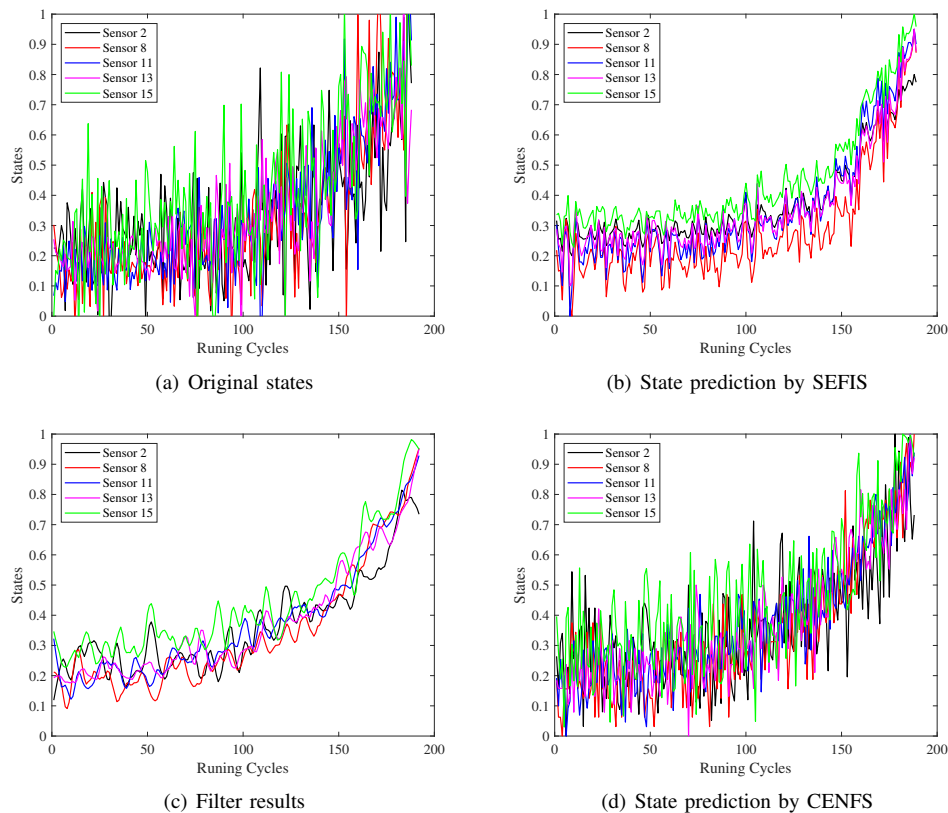
Fig. 7. State prediction for turbofan engine

Fig.7(d) shows the prediction results using CEFNS. In terms of RMSE evaluation metrics, the proposed SEFIS achieves better prediction performances with "RMSE=0.1407" compared to CENFS with "RMSE=0.1809'.

VII. Conclusion

This paper introduces a novel statistical evolving fuzzy inference system (SEFIS) for nonlinear process modeling in the non-Gaussian noise environments. The proposed system utilizes the statistical criterion and is capable of simultaneously constructing the rule components and adapting the rule parameters. The system generates a new rule based on the statistical model sufficiency which gives so insight into whether models are reliable and their approximations can be trusted. The statistical model sufficiency would help to achieve an optimal solution either in the expressibility or in the computational burden. An adaptive maximum correntropy Kalman filter algorithm is derived to update the parameters of the evolving rules to cope with the non-Gaussian noises problems, which is robust to noisy data and outliers. The parameter updating process shares an estimate of the uncertainty with the criterion of the structure evolving process to reduce the computation burden. The proposed system is further evaluated using some benchmark problems. Their results suggest that the proposed SEFIS owns faster convergence and better learning accuracy with more compact structure than the RMCEFS and CEFNS although they can resist the non-Gaussian noises. Besides, the results demonstrate that the proposed SEFIS

has better performance than other popular EFSs due to its statistical rule evolution way in both noise-free and non-Gaussian noise environments. This implies that the proposed SEFIS is an effective modeling method for real-time nonlinear process with any uncertainty property. The superior non-Gaussian noise rejection ability of this proposed system makes it a promising candidate for online modeling and control of complex processes with various non-Gaussian noises.

References

[1] M. M. Ferdaus, M. Pratama, S. G. Anavatti, and M. A. Garratt, "Palm: An incremental construction of hyperplanes for data stream regression," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 11, pp. 2115–2129, 2019.

[2] S. Samanta, M. Pratama, and S. Sundaram, "A novel spatio-temporal fuzzy inference system (SPATFIS) and its stability analysis," *Information Sciences*, vol. 505, pp. 84–99, 2019.

[3] I. Škrjanc, J. A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, and F. Gomide, "Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A survey," *Information Sciences*, vol. 490, pp. 344–368, 2019.

[4] Y.-Y. Lin, J.-Y. Chang, and C.-T. Lin, "Identification and prediction of dynamic systems using an interactively recurrent self-evolving fuzzy neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 2, pp. 310–321, 2013.

[5] P. Angelov and X. Zhou, "Evolving fuzzy-rule-based classifiers from data streams," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1462–1475, 2008.

[6] G. Leng, T. Martin, McGinnity, and G. Prasad, "An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network," *Fuzzy Sets and Systems*, vol. 150, no. 2, pp. 211–243, 2005.

[7] D. Ge and X.-J. Zeng, "A self-evolving fuzzy system which learns dynamic threshold parameter by itself," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 8, pp. 1625–1637, 2019.

[8] K. Subramanian, S. Suresh, and N. Sundararajan, "A metacognitive neuro-fuzzy inference system (McFIS) for sequential classification problems," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 6, pp. 1080–1095, 2013.

[9] N. Kasabov and Q. Song, "DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 144–154, 2002.

[10] J. de Jesús Rubio, "SOFMLS: Online self-organizing fuzzy modified least-squares network," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 6, pp. 1296–1309, 2009.

[11] N. Kasabov and Q. Song, "FLEXFIS: A robust incremental learning approach for evolving Takagi-Sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1393–1410, 2008.

[12] E. Lima, M. Hell, R. Ballini, and F. Gomide, *Evolving Fuzzy Modeling Using Participatory Learning*. IEEE, 2010, pp. 67–86.

[13] A. Lemos, W. Caminhas, and F. Gomide, "Multivariable gaussian evolving fuzzy modeling system," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 91–104, 2011.

[14] L. Maciel, R. Ballini, and F. Gomide, "Evolving possibilistic fuzzy modeling for realized volatility forecasting with jumps," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 2, pp. 302–314, 2017.

[15] P. Angelov and D. Filev, "An approach to online identification of Takagi-Sugeno fuzzy models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 484–498, 2004.

[16] P. Angelov and D. Filev, "Simpl_eTS: a simplified method for learning evolving Takagi-Sugeno fuzzy models," in *The 14th IEEE International Conference on Fuzzy Systems*, 2005, pp. 1068–1073.

[17] H.-J. Rong, N.Sundararajan, G.-B. Huang, and P.Saratchandran, "Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction," *Fuzzy Sets and Systems*, vol. 157, no. 9, pp. 1260–1275, 2006.

[18] H.-J. Rong, N. Sundararajan, G.-B. Huang, and G.-S. Zhao, "Extended sequential adaptive fuzzy inference system for classification problems," *Evolving Systems*, vol. 2, no. 2, pp. 71–82, 2011.

[19] M. Pratama, S. G. Anavatti, P. P. Angelov, and E. Lughofer, "PANFIS: A novel incremental learning machine," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 1, pp. 55–68, 2014.

[20] R.-J. Bao, H.-J. Rong, P. Angelov, B. Chen, and P. K. Wong, "Correntropy-based evolving fuzzy neural system," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 5, pp. 1324–1338, 2018.

[21] H.-J. Rong, Z.-X. Yang, and P. K. Wong, "Robust and noise-insensitive recursive maximum correntropy-based evolving fuzzy system," *IEEE Transactions on Fuzzy Systems*, 2019.

[22] A. Mukherjee and A. Sengupta, "Likelihood function modeling of particle filter in presence of non-stationary non-gaussian measurement noise," *Signal Processing*, vol. 90, pp. 1873–1885, 2010.

[23] S. Thrun, W. Burgard, and D. Fox, *Probalistic robotics*. Kybernetes, 2006.

[24] W. L. Buntine and A. S. Weigend, "Bayesian back-propagation," *Complex Systems*, vol. 5, pp. 603–643, 1991.

[25] H.-J. Rong, G.-B. Huang, N. Sundararajan, and P. Saratchandran, "Online sequential fuzzy extreme learning machine for function approximation and classification problems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 4, pp. 1067–1072, 2009.

[26] K. B. Cho and B. H. Wang, "Radial basis function based adaptive fuzzy systems and their applications to system identification and prediction," *Fuzzy Sets and Systems*, vol. 83, pp. 325–339, 1996.

[27] K. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.

[28] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[29] J. Zhao and C.-M. Lin, "Wavelet-TSK-type fuzzy cerebellar model neural network for uncertain nonlinear systems," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 3, pp. 549–558, 2019.

[30] W. Jianjun, Y. Hua, Z. Hongmin, and C. Yiting, "Global typhoon path dataset (2018)," *National Tibetan Plateau Data Center*, 2019. [Online]. Available: http://dx.doi.org10.11888/Disas.tpdc.270205

[31] A. Saxena and K. Goebel, "Turbofan engine degradation simulation data set," *NASA Ames Prognostics Data Repository, NASA Ames Research Center*, 2008. [Online]. Available: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/\#turbofan

[32] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.