System Identification of Stochastic Nonlinear Dynamic Systems using Takagi-Sugeno Fuzzy Models

Salman Zaidi

Schriftenreihe Mess- und Regelungstechnik der Universität Kassel

Band 09

mrt
Univ.-Prof. Dr.-Ing. Andreas Kroll

kassel university press

Schriftenreihe Mess- und Regelungstechnik der Universität Kassel

Band 9 / Vol. 9

Herausgegeben von / Edited by
Univ.-Prof. Dr.-Ing. Andreas Kroll

System Identification of Stochastic Nonlinear Dynamic Systems using Takagi-Sugeno Fuzzy Models

Salman Zaidi

mrt

kassel university press

This work has been accepted by the Faculty of Mechanical Engineering of the University of Kassel as a thesis for acquiring the academic degree of Dokter der Ingenieurwissenschaften (Dr.-Ing.).

Supervisor: Prof. Dr.-Ing. Andreas Kroll
Co-Supervisor: Prof. Dr. rer. nat. Bernhard Sick

Defense day: 7th May 2018

Bibliographic information published by Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.dnb.de.

Zugl.: Kassel, Univ., Diss. 2018
ISBN 978-3-7376-0650-9 (print)
ISBN 978-3-7376-0651-6 (e-book)
DOI: http://dx.medra.org/10.19211/KUP9783737606516
URN: https://nbn-resolving.org/urn:nbn:de:0002-406517

www.upress.uni-kassel.de

Printed in Germany

Foreword

This work summarizes my five and half years of research at the department of Measure and Control Engineering, University of Kassel, Germany. First of all, I am very grateful to Prof. Dr.-Ing. Andreas Kroll for showing his confidence in me and to give me a chance to join his research group. His invaluable comments, suggestions, and even critics, guided me towards a successful completion of my dissertation. I would also like to thank Prof. Dr. Bernhard Sick for being my co-supervisor and for his insightful comments and precious time. In addition, I sincerely acknowledge the German Academic Exchange Service (DAAD), for financial and administrative support during the four years duration of my research. I must acknowledge the contribution and role of Dr.rer.nat. Hanns-Jakob Sommer. Besides his in-depth knowledge, he is a great motivator. I feel privileged to learn a lot of things from him. I would further like to extend my gratitude to Prof. Dr.-Ing Kroll for entrusting my skills and involved me in the exciting project of inverse modelling, which not only helped me financially, but also enhanced my knowledge in this fascinating area.

I am very thankful to all my colleagues at the Department of Measurement and Control Engineering for the help that they provided me and specially for the friendly working environment. I would like to especially thank my colleague and friend David Arengas, who has always been a great support for me. I feel honored to mention the names of Dr. Zeeshan Shareef and M.Sc. Abdul Rafay Khatri, for their continuous motivation at each and every step of my PhD.

My last thoughts go to my parents in Pakistan. I thank them for inculcating in me the pursuit of learning and scholarships. Most of all, I am eternally grateful to them for their loving affection and prayers. The integrity of my father and the meticulous hardworking nature of my mother have been a guiding light for me throughout my life. Needless to mention, I would also like to thank my siblings because of their support, especially in the early stage of my education. Although I have been living alone for 6 years because of my higher studies but still I can feel their sincerity and affection. At the end, I owe my deepest thankfulness to my wife, Raisa Zaidi, for her understanding, inspiration and motivation. I thank her for appreciating the value of my work, and for believing in my judgment to pursue this degree despite all the hardship.

Kassel, May 2017 *Salman Zaidi*

to my family and friends

Abstract

This work deals with Takagi-Sugeno (TS) fuzzy modelling of nonlinear dynamic systems with uncertainties. Specifically, some novel approaches to estimate Nonlinear Output Error (NOE) models using TS fuzzy models for a class of nonlinear dynamic systems having variability in their outputs is presented in this dissertation. The variability in the system output may be attributed to inherent stochasticity in system dynamics, external disturbances and noise. Instead of using unrealistic assumptions about uncertainty, the most common of which is normality, the proposed methodology tends to capture effects caused by the real uncertainty observed in the data. The methodology requires that the identification method must be repeated offline a number of times under similar conditions. This leads to multiple input-output time series from the underlying system. These time series are pre-processed using the techniques of statistics and probability theory to generate the envelopes of response at each time instant. Three types of envelopes are proposed in this research: the max-min, percentile, and confidence interval based envelopes. By incorporating interval data in fuzzy modelling and using the theory of symbolic interval-valued data, a TS fuzzy model with interval antecedent and consequent parameters is obtained. The proposed identification algorithm provides for a model for predicting the center-valued response as well as envelopes as the measure of uncertainty in system output. It is demonstrated in three case studies: with artificial data, and with the measured data obtained from an electro-mechanical throttle valve as well as a longitudinal servo-pneumatic drive.

Kurzfassung

Diese Arbeit behandelt die Takagi-Sugeno (TS) Fuzzy-Modellierung von nichtlinearen dynamischen Systemen mit Unsicherheiten. Insbesondere werden in dieser Dissertation einige neuartige Ansätze zur Schätzung von nichtlinearen Ausgangsfehlermodellen (NOE) unter Verwendung von TS-Fuzzy-Modellen für eine Klasse von nichtlinearen dynamischen Systemen mit Variabilität in den Ausgangssignalen vorgestellt. Die Variabilität der Systemausgabe kann der inhärenten Stochastizität der Systemdynamik, externen Störungen und Rauschen zugeschrieben werden. Anstatt unrealistische Annahmen über die Unsicherheit zu treffen, üblicherweise wird eine Normalverteilung angenommen, erfasst die vorgeschlagene Methodik die Auswirkungen der Unsicherheiten, die in realen Messdaten zu beobachten sind. Das vorgestellte Verfahren erfordert, dass die Messung der Ausgangssignale unter ähnlichen Bedingungen mehrmals durchgeführt wird. Dies führt zu mehreren Zeitreihen der Ein- und Ausgangsgrößen vom zugrunde liegenden System. Diese Zeitreihen werden mit Methoden der Statistik und der Wahrscheinlichkeitstheorie vorverarbeitet, um Hüllkurven zu erzeugen. Dabei wurden drei Arten von Hüllkurven vorgeschlagen, und zwar Max-Min-, Perzentil- und Konfidenzintervalle. Aus der Integration von Intervalldaten in die Fuzzy-Modellierung und unter Verwendung der Theorie symbolischer intervallwertiger Daten resultiert ein TS-Fuzzy-Modell mit Intervallprämissen sowie Intervallparametern der lokalen Modelle. Der vorgeschlagene Identifikationsalgorithmus liefert ein Modell zum Vorhersagen der mittleren Systemausgabe sowie der zugehörigen Hüllkurven der Ausgabewerte als Maß für die Unsicherheit. Dieses Verfahren wird anhand von drei Fallstudien demonstriert und zwar mit künstlich erzeugten Daten sowie mit Messwerten einer elektromechanischen Regelklappe sowie eines servopneumatischen Linearantriebs.

Contents

List of Symbols

Symbol	Meaning
α	Significance level
$\mathcal{B}_{\mathsf{r}}$	Borel sigma-algebra on r
$\mathbb{E}$	Expected value
$\mathcal{E}$	i.i.d zero mean and finite variance additive measurement noise
Θ_{GMM}	Parameter space of $\boldsymbol{\theta}_{\text{GMM}}$
$\boldsymbol{\theta}_{\text{GMM}}$	Parameter vector of GMM
$\boldsymbol{\theta}_{\text{MLE}}$	MLE estimate of unknown parameter vector
$\mathcal{L}(\cdot)$	Likelihood function
μ	Mean of Gaussian distribution
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$	Normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ
π_l	l-th mixing coefficient of GMM
Σ	Covariance matrix of Gaussian distribution
σ	Standard deviation of population
$\hat{\sigma}$	Standard deviation of sample
$\mathcal{U}(\mathsf{r}; a, b)$	Uniform distribution of the random variable $\mathsf{r} \in [a, b]$
ϕ	Angular position of the electro-mechanical throttle valve
$\phi_{h,i}$	Phase of the i-th harmonic
$\boldsymbol{\phi}_{\text{GMM}}$	Parameter vector of a mixture in GMM
Ω_{r}	Sample space of r
$\omega_{h,i}$	Angular frequency of the i-th harmonic in rad/s
$a_{h,i}$	Amplitude of the i-th harmonic
$\boldsymbol{a}_i$	Consequent parameters

Symbol	Meaning
$\boldsymbol{a}^{\mathrm{SW}}$	Parameter vector of SW statistic
E	Error function
F	Frequency in Hz
$F_{\max}$	Maximum frequency
$F_{\min}$	Minimum frequency
$F_{\mathsf{r}}(r_p)$	CDF of r evaluated at r_p
$f(\cdot)$	Probability Mass Function (PMF)
$f_{\mathsf{r}}(r)$	PMF of r
h_b	Bin width of histogram
h_w	Window width
$I(\cdot)$	Indicator function
J_{FCM}	Objective function of the fuzzy c-means clustering
$J_{\text{IDB-CV-FCM}}$	Objective function of the Interval-Valued based Crisp-Valued Fuzzy C-Means clustering
$J_{\mathrm{OLS}}^{\mathrm{GBL}}$	Objective function of the ordinary global least squares
$J_{\mathrm{OLS}}^{\mathrm{LCL}}$	Objective function of the ordinary local least squares
K	Index set for k
$Ker(\cdot)$	Kernel function
k_c	Constant in Chebyshev's inequality
M	Number of experiments
$\boldsymbol{m}$	Vector of expected values of the order statistics
N	Length of one experiment
$\mathbb{N}$	Set of natural numbers
N_h	Number of harmonics in a multisine signal
N_{o}	Total number of observations
N_p	Total number of parameters
n_c	Number of data points in the current interval
n_I	Value from the set of natural numbers
n_i	Ideal number of data points in the given interval
n_z	Number of intervals
P_{r}	Probability measure of r
$P_{\mathsf{r}}(a < \mathsf{r} < b)$	Probability that r lies between a and b
$p(\cdot)$	Probability Density Function (PDF)

Symbol	Meaning
p_{ch}	Change probability
$p_l(\cdot)$	l-th density component (mixture)
$p_{\mathsf{r}}(r)$	PDF of r
$\hat{p}_{\mathsf{r}}^{\text{NH}}(r)$	Normalized histogram of r
$\hat{p}_{\mathsf{r}}^{\text{KDE}}(r)$	KDE PDF estimate of r
p-value	Test statistic relative to a statistical model
prc	Percentile
$prc^{(low)}\%$	Lower percentile value in percentage
$prc^{(upp)}\%$	Upper percentile value in percentage
Q	Super population
$Q(\cdot)$	Expected value of log likelihood
Q_l	l-th sub-population
q	Total number of components (mixtures) in GMM
R	A set of values of $\boldsymbol{r}$
$\mathbb{R}$	Set of real numbers
r	Random variable
r	Scalar value of r
$\boldsymbol{r}$	Vector value of r
r_0	Origin of histogram
S	A set of values of $\boldsymbol{s}$
$\boldsymbol{s}$	Vector value of latent or missing data
T_s	Sampling time
t	Continuous value of time
$\boldsymbol{u}$	Deterministic input vector
u_{offset}	Offset of the input signal
u	Scalar value of the input signal
$\boldsymbol{v}_i$	i-th cluster prototype
W	SW test statistic
$\mathbf{x}$	Consequent or regressor variable
Y	Stochastic output matrix
y	Scalar value of the output signal
y_k	Expected value of the stochastic output variable y_k at the k-th instant
$\tilde{Y}$	Interval output matrix (interval form)

Symbol	Meaning
$\tilde{y}$	Output signal in interval form
$\underline{y}$	Lower bound of output
$\bar{y}$	Upper bound of output
$\hat{y}_i$	Output of the i-th rule
Z	Input-output data matrix
$\boldsymbol{z}$	Antecedent or scheduling variable

List of Abbreviations

Symbol	Meaning
AD	Anderson-Darling
AIC	Akaike Information Criterion
AICc	Corrected AIC
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
BIC	Bayes Information Criterion
CDB	Crisp-Data Based
CDF	Cumulative Distribution Function
CI	Confidence Interval
CovInd	Coverage Index
CV	Crisp Valued
CWC	Coverage-Width-based Criterion
DOE	Design of Experiment
EM	Expectation Maximization
FLS	Fuzzy Logic System
FOU	Footprint of Uncertainty
FS	Fuzzy Set
GMM	Gaussian Mixture Model
GT2	General Type-2
HIL	Hardware-in-the-Loop
IDB	Interval-Data Based
IFCM	Fuzzy c-means clustering method for symbolic interval data
INFUMO	INterval FUzzy MOdel

Symbol	Meaning
IT2	Interval Type-2
ITSFM	Interval TS Fuzzy Model
i.i.d.	Independent and identically distributed
IV	Interval Valued
KDE	Kernel Density Estimation
KS	Kolmogorov-Smirnov
LF	Lillifors
LP	Linear Programming
M-SVM	Multi-category Support Vector Machine
MAP	Maximum A Posteriori
MaxAE	Maximum Absolute Error
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MLE	Maximum Likelihood Estimate
MRLP	Multi-category Robust Linear Programming
NARMAX	Nonlinear Autoregressive Moving Average model with eXogenous inputs
NARX	Nonlinear AutoRegressive with eXogenous Inputs
NMSE	Normalized Mean Squared Error
OSAP	One-Step-Ahead Prediction
PDF	Probability Density Function
PFLS	Probabilistic Fuzzy Logic System
PFS	Probabilistic Fuzzy Set
PI	Prediction Interval
PMF	Probability Mass Function
PWARX	Piece-Wise AutoRegressive eXogenous
PWM	Pulse Width Modulated
QAMPRS	Quasi Amplitude Modulated Pseudo Random Signal
RMSE	Root Mean Squared Error
SA	Simulated Anealing
SDA	Symbolic Data Analysis
SFS	Sequential Forward Selection
SI	System Identification
SISO	Single-Input-Single-Output

Symbol	Meaning
SW	Shapiro-Wilk
T1	Type-1
T2	Type-2
TR	Type-Reduced
TS	Takagi-Sugeno

1

Introduction

This chapter starts with the motivation of the work followed by the state-of-the-art. The contribution of this dissertation is highlighted next and a brief discussion is provided at the end.

1.1 Motivation

Modeling of dynamic systems is important in applications such as signal processing and decision analysis, simulation and prediction (forecasting), fault diagnosis, hardware-in-the-loop (HIL) simulation, and model-based control. Three types of modeling approaches can be distinguished, namely white-box, gray-box and black-box modeling [81]. White-box models are based on first principles. Owing to overly complex dynamics, lack of system insight, uncertain or unknown system environment, time and accuracy constraints etc., these models are extremely hard or impossible to obtain in many practical situations. Alternatively, black-box models can be used. They use only experimental data and do not require domain knowledge for their estimation. Gray-box models (also referred to as semi-physical models) are midway between white-box and black-box models [51]. These models combine the qualitative knowledge of the system with the empirical data. Black-box models have the advantage that no physical insight of the system is required, and thus many unrealistic assumptions need not be made. Most System Identification (SI) algorithms belong to this category.

In modeling and control theory, system identification deals with building mathematical models of dynamic systems from measured-input-output data. Being derived from data, their accuracy is highly dependent on the information content and the quality of the data used for identification. The major steps in all SI techniques are experiment design, model postulation, parameter estimation and model validation. A number of model types have been successfully used in SI, for instance artificial neural networks [80, 85], piecewise affine systems [21, 28], Takagi-Sugeno (TS) Fuzzy Logic Systems (FLSs) [7, 52, 108], the Kolmogorov-Gabor polynomial and parametric Volterra-Series models [81], and Nonlinear Autoregressive Moving Average models with eXogenous inputs (NARMAX model) [14], to name a few.

Two major types of FLSs can be distinguished, namely Mamdani [68] and Takagi-Sugeno FLSs. The difference between these two models lies in the consequent structure. In case of a Mamdani fuzzy system, the consequent is a fuzzy set, whereas in a TS fuzzy system, the consequent is a mathematical function. While Mamdani FLSs [1] have widespread acceptance due to their suitability for linguistic concepts or variables, TS FLSs are computationally efficient and better mathematically framed. Moreover, the concepts of linear control theory can be transferred to nonlinear systems using validity functions and local affine models of TS fuzzy systems. Owing to their universal approximation capability, the general Single-Input Single-Output (SISO) TS fuzzy system with linear rule consequents can uniformly approximate any continuous function on a compact set with arbitrarily high precision [113, 122].

The term “certain” means data are given as exact constants and “uncertainty” means lack of certainty and having less knowledge. Due to uncertainty, it is impossible to retrieve the existing state or future outcome [40]. According to [126], “uncertainty is an attribute of information”. Different kinds of uncertainties arise during the modeling process on account of limited physical insight of the system, inherent stochasticity in the system, unrealistic assumptions about the distribution of noise and disturbance (such as Gaussian etc.), unmodelled inputs/disturbances, unmodelled dynamics, limited quantity and quality of available data, and/or limited model complexity for control-oriented modeling. The conventional or Type-1 [2] (T1) TS FLS may not adequately describe the system at hand. In the T1 FLS, once the Fuzzy Sets (FSs) are fixed, all uncertainty disappears because these sets have fixed, certain membership values [71]. Consequently, T1 FLSs do not have the ability to effectively handle [3] uncertainties. In addition, they have deterministic rule base and fixed consequent

[1] Mamdani FLSs are particularly suited for fuzzy expert system type applications, where the rule base is often collected from people, which renders them more transparent, i.e. easier to interpret.

[2] Being consistent with [71], the terms type-1 and type-2 have been used in this dissertation in order to distinguish the two types of FLSs/FSs.

[3] By handling uncertainty, it is meant to model and minimize the effect of it.

structure, which limits their ability to handle uncertainties. Classically, these T1 FLSs are estimated from a crisp input-output data set, which represents no uncertainty in input and output, and thus a crisp set of antecedent and consequent parameters are obtained. Also, the output of a T1 FLS is a crisp number without any uncertainty description associated with the output.

In most cases of SI using traditional TS FLSs, only the point prediction of output values (i.e. crisp or point-valued based output values) is provided. These values are without any quality tag or confidence, which help determine how good the point prediction is or how much the output value can deviate from the point prediction. If there are uncertainties in the system dynamics, e.g. because of friction, it may be preferable to obtain several data sets, each having the same input signal and initial conditions. Consequently, such data sets can be used to capture the variability in system dynamics. On the contrary, only a single identification experiment is usually performed, which does not take into account of the variation of the output time series in different data sets. Variability in the output time series is defined as the inherent variation of the output time series when the same input signal is applied to the system under the same initial conditions. Stochastic systems show variability in their output due to their inherent stochastic nature. If such a stochastic system is to be modeled using input-output data, it would be advantageous to capture output variability in different data sets obtained by repeating the identification experiment multiple times, keeping the input and initial conditions constant. Thus, when dealing with stochastic systems, using only one data set for identification ignores the variability of the system output and the model is developed only for one data set, which represents one realization of the output time series. Thus for reliable identification, it is necessary to have information on the variability of the output besides a predicted value of the output. Dependent on the character of the variability, this information could be max/min boundaries of the output or statistical information on its spreading. Such a model should be able to provide a range of output values where the predicted output is expected to lie in the presence of all these uncertainties. Moreover, the model should have uncertain parameters to show the variations in prediction, which can be used for robust control.

Finally, it is desirable to develop a model for the Nonlinear Output Error (NOE) case, also referred to as parallel configuration or recursive evaluation. The NOE model is important for many practical reasons, e.g. accurate simulation, model predictive control. The NARX and NOE models are described in detail in Sec. 4.3.

It is worth mentioning here that the estimated model should reflect uncertainties as realistically as possible. In most of the cases, a normality assumption is made, for instance

using Gaussian Processes (GPs). However, this contradicts the reality in many practical nonlinear systems, such as the electro-mechanical throttle, where the observed stochastic behavior is mainly due to friction. Fig. 1.1 shows the resulting output time series when a multisine input is applied to the electro-mechanical throttle multiple times, see Sec. 3.6.2 for details. Due to the stochastic behavior of friction, a band of output time series is observed.

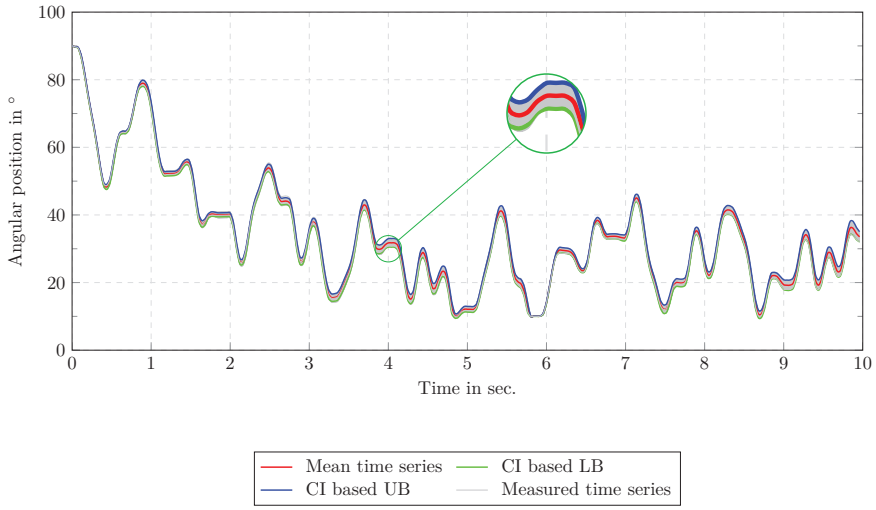


Figure 1.1: The output signal (the angular position of the throttle) for $M = 80$ experiments

By examining the distribution of output values at each time instant, it is observed that the output values are distributed non-normally and unsymmetrically, as illustrated in Fig. 1.2. Modelling such a system using GP or other probabilistic methods which require the distribution to be normal (the Gaussian Mixture Model (GMM) is excluded) would lead to erroneous results. As it is computationally infeasible to process all these histograms at every time step, it is desired to first transform this data into computationally feasible form by using statistical techniques, for instance, maximum-minimum values, confidence bounds or percentiles.

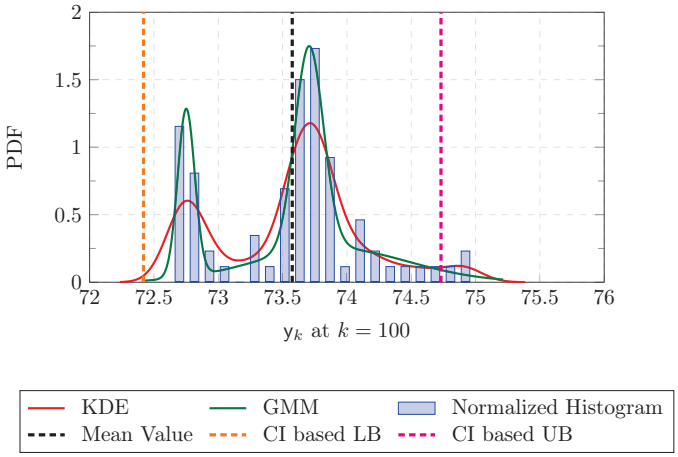


Figure 1.2: Uncertainty analysis of y_k at $k = 100$

1.2 State-of-the-Art

The application of the classical T1 FLS is limited to the cases where the mathematical model of a system is described by T1 fuzzy sets which have no uncertainty about the membership function values. Modeling a stochastic system having stochastic rule base or interval/ histogram output requires special treatment.

Zadeh first introduced the concept of fuzzy sets in his seminal work in 1965 [127]. Soon after that, he proposed the theory of fuzzy logic in 1973 [123]. Although, the original concept of T2 FLS was presented in 1985 by Zadeh in his work [124], it did not gain much popularity until the work of Mendel in 1999 [43]. Since then, the T2 FLS [71, 88] has gained a lot of popularity in the fuzzy logic research community. The difference between the T1 FLS and T2 FLS lies in the type of fuzzy sets used in the fuzzy inference mechanism. T1 FLSs use T1 FS in which for each given value of the linguistic or fuzzy variable in the universe of discourse, the membership of the variable takes a crisp value between 0 and 1, i.e. in the close set $[0, 1]$. Whereas, in T2 FLSs, which use T2 FSs, the membership functions are themselves fuzzy, leading to primary and their corresponding secondary membership functions. It is the additional degree of freedom in T2 FLSs that makes it possible to better model uncertainties. In [73], it is mentioned that there exist at least four sources of uncertainties in T1 FLSs justifying the use of T2 FLSs:

1. The meanings of the words that are used in the antecedents and consequents of rules can be uncertain (words mean different things to different people).

2. Consequents may have a histogram of values associated with them, especially when knowledge is extracted from a group of experts who do not all agree.

3. Measurements that activate a T1 FLS may be noisy and therefore uncertain.

4. The data that are used to tune the parameters of a T1 FLS may also be noisy.

According to Mendel and John, "Type-2 fuzzy sets allow for linguistic grades of membership, thus assisting in knowledge representation, and they also offer improvement on inference with type-1 fuzzy sets [73]". Since all these uncertainties can be translated into uncertainties in fuzzy set membership functions, Mendel and his co-authors [42, 44, 45, 60, 71] motivated the use of Type-2 (T2) FLSs in cases where classical T1 FLSs are not capable of handling the higher degree of uncertainty adequately. According to Mendel, by handling uncertainty, it is meant to model and minimize the effects of it [71]. Thanks to their additional design degrees of freedom, T2 FLSs accomplish this task, leading to more design parameters, which are determined during the training stage, resulting in increased robustness and higher accuracy in terms of error reduction as compared to their counterpart T1 FLSs especially when the noise is heterogeneous (noise variance changing with time) and has relatively large value of variance.

T2 fuzzy logic is very useful when it is difficult to determine the exact membership functions of the fuzzy sets [73]. Since T2 FLSs are computationally expensive and harder to understand, their use is still limited, especially in real-time applications, and there has to be a solid justification behind their use. For instance, Mendel suggested to use the T2 FLS in applications such as function approximation, forecasting, learning linguistic membership grades, and decision making etc., when the measurement noise is non-stationary and the mathematical description of the nonstationarity is unknown [71]. To limit the tremendous computational complexity associated with General Type-2 (GT2) FLSs, some simplified variations have been proposed, namely Interval Type-2 (IT2) FLSs [59], Quasi-Type-2 (QT2) FLSs [74], α-plane representation based T2 FLSs [75], and z-slices based T2 FLSs [112]. T2 FLSs have shown superior performance in terms of robustness and error reduction in the presence of large uncertainties as compared to their counterpart T1 FLSs in various applications, such as modeling and identification [2, 36, 37, 39, 115, 118], control of mobile robots [30, 117, 119], decision making [77, 78], time series forecasting [44, 76], stock price forecasting [82], motion planning [8, 86], image processing [16, 136], equalization of nonlinear fading channels [61, 62], to name a few. For a concise review of the industrial

applications of T2 FLS, see [24]. Nie et al. [83] analyzed the modeling capability of the T1 FS and IT2 FS and concluded that IT2 FLSs achieve higher modeling accuracy on account of the unique characteristics of the IT2 FS, i.e. the additional design freedom of IT2 makes it robust against the variation of its Membership Function (MF) from the ideal one (because of the presence of uncertainties) in terms of its centroid deviation from the ideal one giving the IT2 FLS the potential to achieve higher modeling accuracy as compared to the T1 FLS. One of the main difficulties in T2 FLSs is the determination of T2 FSs. Zarandi et al. [135] used the T2 FLS to model desulphurization process of a real steel industry in Canada, in which they used fuzzy c-means clustering and Gaussian Mixture Model (GMM) to systematically estimate the primary and secondary membership functions, respectively. Wang et al. [115] used T2 TS fuzzy model for modeling dynamic systems having measurement noise as uncertainty. They used the improved nearest-neighbourhood clustering [114]. However, these approaches only judged the performance of T2 FLSs based on their higher accuracy as compared with their counterparts T1 FLSs and ignored one of the key advantages of using T2 FLSs over T1 FLSs, i.e. their ability to provide a Type-Reduced (TR)[4] set as a measure of quantitative uncertainty associated with the output on account of various uncertainties encountered during the modeling process. One of the studies on the TR set as a measure of input-output uncertainty was carried out in [65]. Keeping note of the fact that a T2 FS can be considered to be composed of a number of embedded T1 FS, the authors demonstrated that the output TR set over estimates the output uncertainty range when compared to the results of the Monte Carlo simulation. In summary, T2 FLSs have been generally renowned for their capability to model various sources of uncertainties. One of the key advantages of T2 FLSs is that not only do they provide a crisp output at the end, but also an uncertainty description as additional information, in terms of a type reduced set. The main drawback is the increased computational complexity, especially in type reduction, and the lack of a systematic design approach to model effectively uncertainties in the secondary membership function of T2 FSs. Furthermore, the accuracy of the TR set in terms of actual fuzziness associated with the output has not been well established yet in terms of modeling uncertainty, and to date there is no systematic way to incorporate and meaningfully quantify data uncertainty in the inference mechanism of T2 FLSs.

Two types of uncertainties should be distinguished, namely statistical or probabilistic uncertainty (more commonly known as randomness) and non-statistical or fuzzy uncertainty (also called as fuzziness). The fuzziness and randomness are two conceptually different kinds of uncertainties and should be handled with the fuzzy logic and the probability theory, respectively. The fuzzy logic and probability theory are complementary rather then competitive

[4]TR is a T1 FS that provides a measure of uncertainty about the crisp output of a T2 FLS.

[98, 125]. Meghdadi [70] proposed probabilistic fuzzy logic and probabilistic fuzzy systems, in which he combined the non-statistical and statistical uncertainty, in the paradigm of fuzzy logic and probability theory, respectively, in a unique and unified framework. Although, owing to their universal approximation capability to approximate any real continuous function on a compact set to arbitrary accuracy [113], fuzzy sets have been attributed to successfully model complex deterministic systems, their performance degrades significantly by the existence of randomness or probabilistic elements [70]. In the proposed probabilistic fuzzy system, both the probability theory and the fuzzy theory co-exist in a combined framework to better represent non-deterministic real world systems. Moreover, the conventional fuzzy system can be regarded as a special case of the probability fuzzy system with zero degree of randomness. Notwithstanding the research conducted on the probabilistic fuzzy system [70], random fuzzy variable [19], and probabilistic interpretation of fuzzy sets [58] etc. describing the relationship between randomness and fuzziness, they cannot be directly applied to engineering applications [66]. Motivated by the T2 FLS, Liu et al. introduced a special type of T2 FLS that was based on the probability theory, the Probabilistic Fuzzy Logic System (PFLS) [66, 67]. The PFLS is also referred to as the three-dimensional FLS [57]. Zhang et al. [137] developed an efficient configuration for the PFLS by unifying the probabilistic inference and the fuzzy inference based on the continuous form of the probabilistic fuzzy set. The PFLS is based on Probabilistic Fuzzy Sets (PFSs), which are different from T2FSs in that there is a Probability Density Function (PDF) in the PFS, instead of a fuzzy set as in the T2 FS, in the third dimension of the fuzzy set. Different from the T2 FS, The membership grade of PFS is a random variable with a certain PDF. The inference mechanism in PFLS is completely based on the probability theory. The PFLS has the capability of modeling a system with both non-stochastic (or non-statistical) and stochastic (or statistical) uncertainties thanks to the presence of the primary fuzzy MF in the second dimension, and the PDF in the third dimension. The output of the PFLS is a random variable with a certain PDF which actually provides a measure of stochastic uncertainty associated with the output. The main limitations of PFLS include computational complexity, lack of systematic way of determining the PDF from the data, and the interpretation of the output PDF as the observed PDF of the output. Furthermore, the PDF in the second dimension of T2 FS is often assumed to be a Gaussian distribution, which makes the modeling inaccurate when the underlying distribution is non-Gaussian.

Classical fuzzy models only provide a point or average prediction[5] with no description of the reliability of the prediction. However, it is often required to have the information of spread, for example in robust control. Inspired by the fact that the model response should

[5]The predicted output is expressed by a crisp or a point-valued real number

be associated with the quality tag or confidence measure, Khosravi et al. [47] proposed the Prediction Interval (PI) based Adaptive NeuroFuzzy Inference System (ANFIS). A PI represents the uncertainty in prediction and offers a range that the target is highly likely to lie within [47]. A number of methods dealing with PI construction have been developed for Artificial Neural Networks (ANNs) [35, 48, 84, 110], but such methods were not applied to fuzzy systems until the work [47]. The author applied the delta method to construct the prediction interval based on the additive noise term and the model variance. In addition, the author defined a criterion namely the Coverage-Width-based Criterion (CWC). This criterion was based on both the Prediction Interval Coverage Probability (PICP), which determines the validity of PIs (the predicted output lies within the PI), and the Normalized Mean Prediction Interval Width (NMPIW), which describes how informative a PI is (the width of the PI). They used simulated annealing for optimizing the CWC. One of the main limitations was regarding the assumption of uncertainty. It was assumed that the combined uncertainty due to noise and modal variance (based on the Jacobian function of the model parameters) is independent and identically distributed (i.i.d.) and Gaussian. Moreover, the model has no ability to incorporate the inherent variability of the system. In order to relax the normality assumption, the same first two authors introduced a prediction interval based T2 FLS [46, 49]. In that model, the constructed PIs were non-parametric in nature and the model did not make any assumptions about uncertainty distribution. The lower and upper bounds of a PI were taken as the left and right end points of the TR set of the IT2 FLS. As used previously, they used the CWC criterion, which included both the validity and informativeness aspect of the PIs, to measure the quality of the constructed PIs. The CWC based non-continuous and non-differentiable objective function was optimized by the Simulated Annealing (SA) meta-heuristic optimization method. The major disadvantages include very high computational complexity and the model has no ability to incorporate the inherent stochasticity in the system dynamics.

The methods discussed so far have been used for the case of crisp data sets, i.e. data sets which have crisp inputs and outputs. On the contrary, the real data set from the process or plant very often cannot be pinned down to crisp-valued data set owing to the presence of different uncertainties in the real world. With the advent of advancements in the field of data storage and processing of large data sets, the use of complex data sets instead of simple crisp data sets in order to improve the information content of the collected data has emerged in the last few decades. In particular, complex data can be represented and processed in the form of intervals or histograms in the paradigm of Symbolic Data Analysis (SDA) [12, 15]. Fuzzy modeling of dynamic systems using interval data have been rarely discussed in the literature. In 2005, Škrjanc [105, 106] proposed a new type of fuzzy model, referred

to as INterval FUzzy MOdel (INFUMO). INFUMO was particularly useful in cases where a system was defined by a family of uncertain nonlinear functions, or alternatively, when a system has uncertain physical parameters. In INFUMO, an interval l_∞-norm function approximation methodology was developed by using the min-max approach of the Linear Programming (LP) technique and the TS FLS approach. One of the key assumptions for INFUMO was that the upper and lower bounds of response were available as the reference or target time series for the upper or lower fuzzy models. In their considered simulation example of simplified car dynamics as a nonlinear time-invariant system with uncertain physical parameters [3], they obtained the upper and lower bounds of the response by first considering all possible extreme variations of the parameters. It leads to a family of functions, and then the minimum and maximum functions out of that family were determined in order to generate an interval response. The LP technique was applied independently on these bounds to give rise to a fuzzy model with a set of lower and upper parameters. The antecedent structure was determined in the classical way and they assumed that the antecedent variable had no uncertainty. Therefore, one of the shortcomings of INFUMO was its requirement to have crisp antecedent variables, which restricted its usage mainly to static systems, as a result of which the model had consequent parameters in the interval form, whereas the antecedent parameters were crisp. Moreover, since the upper and lower fuzzy models were independent of each other, the approach is not appropriate for interval dynamic systems. The model proposed by Xu and Sun [107, 120], referred to as Interval TS Fuzzy Model (ITSFM), overcame this limitation of INFUMO. The ITSFM has the advantage that it handles the case of crisp input and interval output in a unified framework using interval arithmetic. The obtained consequent parameters were in the form of an interval, whereas for partitioning the input space, the center and half range values of the antecedent variables were clustered separately. The Triangular-norm (T-norm) operator was used to combine the two membership grades. They only considered the case of OSAP or NARX. Furthermore, the model had the limited capability of representing uncertainty in parameters, since only the consequent parameters were represented as intervals. Since the antecedent structure was not handled using the theory of interval clustering, the model was hard to interpret and compared with other commonly known model descriptions.

Alternatively, for handling uncertain information, fuzzy regression analysis [25, 109] can be used; the case is simplified where a model has interval parameters as the special case of fuzzy parameters in [34]. However, it is extremely difficult to determine and interpret the possibility distribution for parameters and/or data in SI of real world dynamic systems, and it is not well defined yet.

1.3 Contribution of the Dissertation

The major contributions of this dissertation lie in the development of novel Crisp- and Interval-Based Data Type-1 FLSs for modeling of stochastic nonlinear dynamic systems. Furthermore, the brief introduction and conceptual sketches for Interval Type-2 FLSs have been presented. Furthermore, uncertainty analysis using statistics and probability theory has been performed to determine the underlying PDF and carry out normality tests. For the developed models, statistics and probability theory based pre-processing techniques have been developed to generate interval data. In these models, Fuzzy clustering and weighted interval linear regression have been used to estimate antecedent and consequent parameters using the theory of symbolic data analysis. The models have been developed for both the NARX and NOE cases. Finally, all these methods have been practically demonstrated on two experimental systems and one academic example.

The developed models extend the multidimensional reference fuzzy sets at the antecedent parts [52] for both crisp and interval data. For the first time, it enhances the classical fuzzy modeling by endowing it with the stochastic theory and the Symbolic Data Analysis in a unique way, thereby giving the model the ability to incorporate variability in system dynamics. Motivated by the fact that a stochastic system produces different time series each time when the same input signal is applied to it under the same conditions, the output variability is captured in the system by repeating the identification experiment a number of times and record multiple data sets for identification and validation. Having obtained the multiple data sets, the multiple output time series are visually and numerically analyzed at each time instant in order to determine several information about uncertainty, such as the histogram of data, the type of distribution (normal/non-normal, symmetric/asymmetric etc.), the moments of distribution, etc. This step is called the pre-processing of data before identification. This pre-processing is essential in order to reduce the data dimensionality, and thus be able to identify a model using the reduced dimensional data. As a result of this pre-processing, the multitude of output time series reduces to three time series, namely, the mean/median/mode or most likely time series for point prediction, and the two time series (lower and upper time series) that provide the envelopes of the response. Three kinds of envelopes are proposed in this research, namely

1. The max-min envelopes
2. The percentile based envelopes and
3. The Confidence Interval (CI) based envelopes.

The idea of using the sample-wise max-min hard bounds was first introduced as envelopes and referred to as max-min envelopes [128]. The main drawbacks of using max-min envelopes include sheer conservativeness, dependability on the given sample, least robust statistics in the presence of outliers, and the fact that a minimum or maximum value of an infinite tail distribution may not be defined. On the flip side, max-min bounds provide for the worst-case variability, which can be of, depending upon the application, practical relevance. The method was successfully demonstrated for the electro-mechanical throttle. Besides envelopes, the method had the drawback that it uses two independent TS fuzzy models for estimating the lower and upper bound output time series, giving rise to lower and upper fuzzy models or the fuzzy model with lower and upper set of parameters. Later on, the concept of probabilistic bounds based on the $(1-\alpha)100\%$ confidence level was introduced (α is called the "significance level" of the underlying distribution) CI [130]. The approach taken in [130] makes the assumption that the expected value is the best point estimator for the given data set. Recall that the expected value is the weighted average or arithmetic mean of a certain PDF. For capturing the spread of the distribution, the CI based envelopes are obtained using the extended Chebyshev's inequality[6] for finite sample size [18, 41]. The key advantages of using this inequality is that it does not require any data insight and thus it makes no assumption about the distribution. Nevertheless, this makes the obtained envelopes extremely conservative. If the distribution of the data is known a priori, then this conservatism can be avoided by using the coverage factor of the underlying distribution, rather than of the 'general' distribution as provided by the extended Chebyshev's Inequality. As previously, two independent TS fuzzy models are estimated individually for estimating the lower and upper bounds or the envelopes of response for the NOE case. Since the structure of Chebyshev's Inequality makes the inherent assumption that the values are symmetrically distributed around the mean, the expected response is obtained by calculating the sample wise average of the response of these two fuzzy models. However, this will in general not be the true expected response. In the work presented in [131], as done previously, the data is pre-processed using the extended Chebyshev's Inequality to generate interval output, and then that data was directly used in the TS modeling, by utilizing the techniques for Symbolic Data Analysis (SDA), and developed a single TS fuzzy model with interval antecedent and consequent parameters. The technique is successfully demonstrated on a servo-pneumatic longitudinal drive in [133]. The developed model is referred to as Interval-Data Based (IDB) Type-1 Takagi-Sugeno FLS (IDB T1 TS FLS).

[6] The inequality is valid for both symmetric and asymmetric distributions. However, less (more) conservative bounds are obtained when the distribution is symmetric (asymmetric).

The IDB T1 TS FLS directly uses the interval data for clustering and it is based on SDA. The aim of SDA is to provide suitable methods (clustering, factorial techniques, decision trees, etc.) for managing aggregated data described by multi-valued variables, where the cells of the data table contain sets of categories, intervals, or histograms [13, 15]. The IDB T1 TS FLS takes interval data and furnishes a crisp partition using the Interval-Data Based Crisp-Valued Fuzzy C-Means (IDB CV FCM) clustering method [1]. The Fuzzy c-means clustering method for symbolic interval data (IFCM) works on the principle of optimizing an adequacy criterion based on suitable squared Euclidean distances between vectors of intervals [1]. The IFCM uses the same Euclidean distance norm for every cluster which results in spherical clusters. On the other hand, if clusters of different shapes and sizes are required, or in other words, if clusters need to be adjusted according to the spread/distribution of data, the fuzzy c-means clustering method for symbolic interval data based on an adaptive squared Euclidean distance between vectors of intervals (IFCMADC) can be used [1]. As a consequence of this type of clustering, the resulting cluster prototypes are in the form of intervals, which defines the antecedent structure of IDB T1 TS FLS. Having done the partitioning, the consequent structure has to be determined. The center and range method suitable for the symbolic interval data [64]. Needless to say, any method for weighted interval regression can be used for estimating the consequent or local model parameters, where weights are actually the membership function values. For keeping the task simple and computationally efficient, the center and range method is chosen.

In all of these models, the model is first estimated for the Nonlinear AutoRegressive with eXogenous Inputs (NARX) model, because it is simpler to estimate as only the past measurements up to the present are used for model estimation and the future response is determined based on these measured values of input and output. A nonlinear optimization algorithm is used to estimate the NOE model, in which the parameters of the NARX model are used as initial values.

The motivation behind choosing the experimental test stands in this research is that the open-loop data showed significant variability in the output especially at low frequencies. The most important and significant common factor of stochasticity in these systems is complex frictional effects (stick-slip effect, static and dynamic friction, Coulomb and Stribeck friction). Beside friction, other factors that can contribute to stochasticity include random spring characteristic and manufacturing imperfections. Besides the experimental test stands of electro-mechanical throttle and servo-pneumatic linear drive, the developed model is demonstrated and validated on a modified simulation benchmark in the work presented in [132]. The chosen system is a second order Single-Input Single-Output (SISO) nonlinear

system [79] with artificially added noise of known distribution at each time instant. This simulation case study is particularly useful because the true time series is known along with its spread and thus the modeling performance can be easily validated. In each of these case studies, a family (band) of output time series is obtained due to the inherent stochasticity in the system as a result of experiment repetition, which forms the basis for the developed identification approach.

The developed models are assessed based on the model complexity, the modeling performance, and the quality of prediction or the spread information.

1.4 Outline of the Dissertation

The rest of this dissertation is organized as follows: Chap. 2 provides the general and mathematical description methods of selected case studies, together with the details of the Design of Experiment (DOE). Chap. 3 presents some uncertainty analysis and description which are required to convert the data in the form that can be used for modeling. Specifically, some techniques have been mentioned that are used for visually inspecting and estimating the PDF that describes uncertainty. Next, the chapter describes different types of envelopes based on the statistics and probability theory. The methods described in this chapter have been applied on the two test stands and the academic example. In Chap. 4, a brief introduction to T2 FLS is provided along with a general overview of the developed modeling approaches based on the crisp and interval data and membership values. The modeling notations used extensively in this dissertation are introduced and tabulated. This chapter ends with the mathematical and visual comparison of the presented models. Chap. 5 and 6 form the crux of the conducted research. Having provided the related work and solid background, these chapters present modeling of the test stands using crisp and interval valued data. Finally, this dissertation ends with some conclusions and outlook on future research in Chap. 7.

2

General Description of the Case Studies

As mentioned in the previous chapter, three case studies have been used in this research to demonstrate the validity and effectiveness of the developed uncertainty modeling and identification approaches. In this chapter, a brief introduction to these case studies will be given along with the motivation behind choosing them. Two of these case studies consist of experimental test stands available in the laboratory of the Department of Measurement and Control Engineering (MRT), Institute for System Analytics and Control (ISAC), University of Kassel, Germany. The simulation example is taken from one of the simulation benchmark systems for identification. A brief description of each of these case studies is given in the following sections.

2.1 Academic Example

2.1.1 Motivation

The system is one of the academic simulation benchmark systems for nonlinear system identification [53, 79]. Since, the developed modeling approaches require the system to be stochastic in nature, this system is made stochastic by artificially introducing an uncertainty of known type by adding noise to the output at each time instant as described in Sec. 3.6.1. The rationale behind choosing this example is that the most likely response of the system is

known as well as the type of uncertainty, and thus the modeling results using the developed modeling framework can be verified as will be shown in later chapters.

2.1.2 Mathematical Description

The original system is a Single-Input Single-Output (SISO) deterministic nonlinear dynamic system described by a second order difference equation as follows:

$$y(k+1) = \frac{y(k)y(k-1)(y(k)+2.5)}{1+y^2(k)+y^2(k-1)} + u(k), \tag{2.1}$$

where k denotes the discrete time index, $u(k)$ and $y(k)$ represent the input and output signal at the k-th time instant, respectively. Since this chapter deals only with the deterministic system description, the system presented here is purely deterministic, i.e. without noise/disturbance/variability etc. Stochasticity is introduced in the system in Chap. 3.

2.1.3 Design of Experiment

A sequence of uniformly distributed random numbers in the closed interval $[-2, 2]$ is chosen as the input signal. A total of $N = 2000$ samples are used for identification and validation. The first half of the data is used for identification and the remaining half is used for validation. The hold time and the change probability are chosen as $N_{\text{hold}} = 10$ samples and $p_{\text{ch}} = 0.5$, respectively, in this case study. Therefore, there is an equal chance whether the input signal remains at its previous value or changes to a new value, as follows:

$$u(k) = \begin{cases} \mathsf{u} \sim \mathcal{U}(\mathsf{u}; -2, 2) & \text{with probability } p = p_{\text{ch}} \\ u(k-1) & \text{with probability } p = 1 - p_{\text{ch}} \end{cases} \tag{2.2}$$

where $\mathcal{U}(\mathsf{u}; -2, 2)$ denotes the uniform distribution of the random variable $\mathsf{u} \in [-2, 2]$. Since the system given by Eq. (2.1) is of second order, its simulation requires two initial conditions, which are taken as $y(0) = 0$ and $y(-1) = 0$ in this research. The resulting input and the output signal used in the case study are shown in Fig. 2.1.

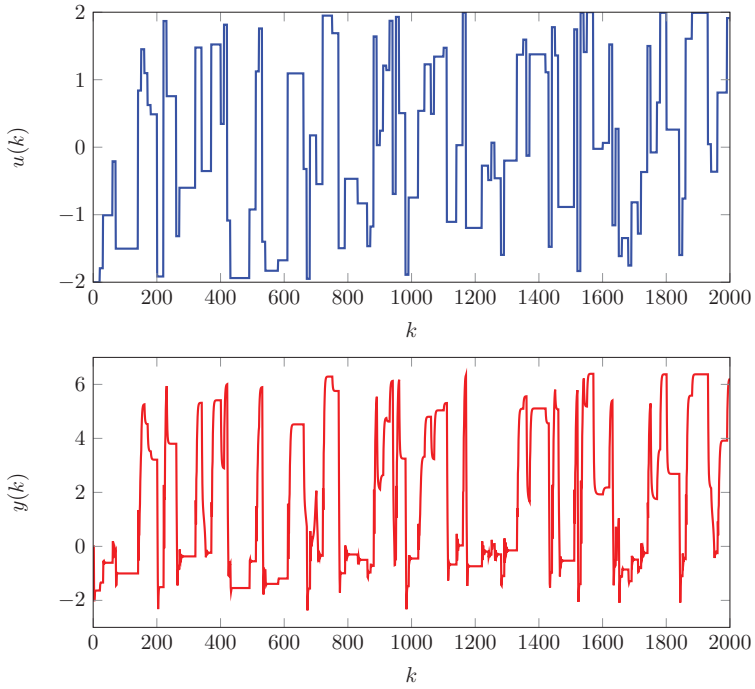


Figure 2.1: The input and the output signal without noise for the simulation case study

2.2 Electro-mechanical Throttle Valve

2.2.1 Motivation

The electro-mechanical throttle was presented as a benchmark problem for nonlinear system identification in [129]. An electro-mechanical throttle is one of the most important mechatronic elements of the automatic engine management system used in modern vehicles. Being standard components in Diesel and Otto combustion engines, electro-mechanical throttles are widely used. For the same reason, they are constructed in simple and cost-effective ways in order to meet robustness and low cost requirements. When modelling an electro-mechanical throttle valve, several complexities are encountered, owing to factors such as manufacturing imperfections, significant friction and nonlinear return spring characteristics.

Physical or white-box models for the electro-mechanical throttle have already been derived in literature, e.g. [94, 101] to name a few. However, such models are not accurate in practice on account of unrealistic assumptions and over-simplification. The randomness in the electro-mechanical throttle is mainly attributed to friction. Being stochastic in nature, friction is often state-dependent and causes complex effects such as stick-slip, see [87] and the references therein for more detail. Besides friction, the factors such as manufacturing defects and imperfections increase the modelling complexity. In addition, some mechanical parts, especially those made up of plastic, may undergo deformation or wear and tear with the passage of time. Additionally, not every physical behavior can be modeled perfectly, for instance, the nonlinear characteristics of return spring and inertia are not typically known. System identification enables one to effectively and efficiently estimate a plausible mathematical model that best describes the underlying nonlinear dynamics. It uses an input-output data set which incorporates all complex effects, and through that data set, it estimates the best model from a selected model class that best describes the given data set with respect to the chosen criteria. One of such attempts consists of the work done by Vasal et al. [111]. They proposed and developed a PieceWise AutoRegressive eXogenous (PWARX) model based on K-means clustering and Multi-category Robust Linear Programming (MRLP). Lebbal et al. [54] developed a model in which they considered various frictional effects in different operating regions of a throttle. In their work, they treated nonlinearities as additive unknown inputs and used the techniques of classification and linear identification to build their model. By utilizing K-means clustering and Multi-category Support Vector Machine (M-SVM), a computationally efficient PWA model was developed by Ren et al. [95]. In that work, a NARX model was estimated and then succeedingly optimized to become a NOE model using the simplex algorithm.

2.2.2 Description of the Test Stand

This system is a SISO nonlinear stochastic dynamic system. The major parts of an electro-mechanical throttle valve include a DC motor, a gear box, a return spring and a throttle plate integrated in metal housing as illustrated in the technology scheme of the throttle in Fig. 2.2. The angular position of the throttle plate is considered as the output signal $y(k)$. The values of $\phi = 10°$ and $\phi = 90°$ correspond to the fully closed and open positions, respectively, of the valve. A Pulse Width Modulated (PWM) signal is used to drive the

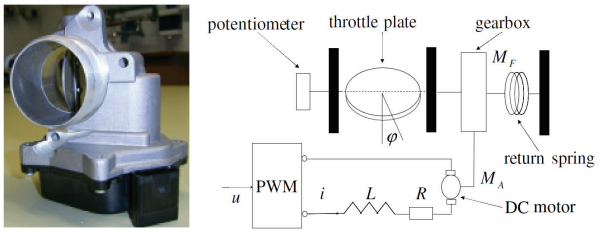


Figure 2.2: The technology scheme of an electro-mechanical throttle

motor. The effective value of this PWM signal can be varied by changing its duty cycle[1], which is taken as the input signal ($u(k)$) measured in percentage. The test stand that has been used in this research for collecting data for identification and validation is illustrated in Fig. 2.3. As can be seen from the figure, the standard automotive throttle is used without load in the laboratory setup.

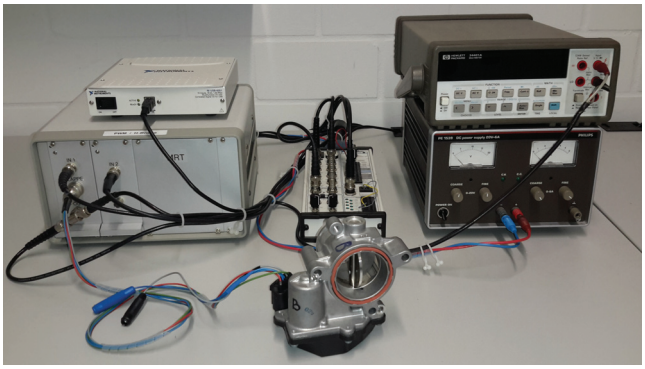


Figure 2.3: The complete test stand of the electro-mechanical throttle valve

2.2.3 Design of Experiment

The test signal used for the identification of the throttle was taken from Ren et al. [95, 97]. The measurements are recorded and filtered at the sampling time of T_s = 10 ms. The sampling time is selected based on a priori knowledge of the system. On one hand, it should

[1] Duty cycle is defined as the ratio between the pulse width (T_{on}) and the period ($T_{on}+T_{off}$) of a rectangular waveform, i.e. Duty Cycle = $\frac{T_{on}}{T_{on}+T_{off}}$, where T_{on} and T_{off} denote the on and off time, respectively. The positive and negative values indicates the polarity of the DC voltage applied to the throttle.

not be too small compared to the underlying dynamics; on the other hand, it should not be too large that it may skip the important high frequency characteristics of the system. Selecting too small a sampling time can cause some serious problems, such as numerical-ill conditioning at the stage of parameter estimation, and high sensitivity to high frequency noise and disturbances etc. Therefore, it is always a trade-off when selecting a sampling time or interval. Roughly, it was selected as 10 % of the settling time of the open loop step response of the system [38, 95].

The input-output data set for this system as a benchmark problem for nonlinear system identification was presented in the work [129]. This data set is now available at [29]. It contains a multisine signal of length $N_{\text{ident}} = 10000$ samples for identification and a skyline signal of length $N_{\text{val}} = 2500$ samples for validation of the model. These signals are shown in Fig. 2.4 and Fig. 2.5, respectively.

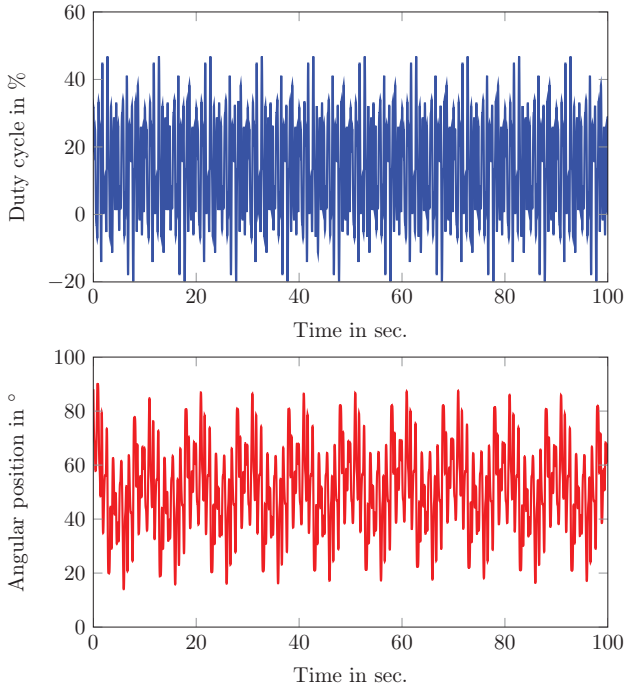


Figure 2.4: Multisine input and the corresponding output signal for identification

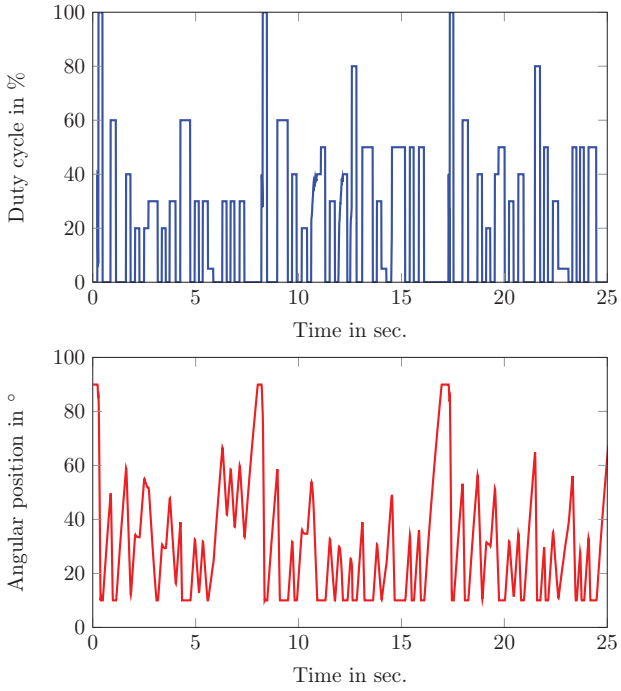


Figure 2.5: The skyline input and the corresponding output signal for validation

The multisine signal has been widely used for nonlinear system identification. One of the key advantages of a multisine signal is in its flexibility of design in order to excite the system at the desired amplitudes and frequencies. Since the experiment is performed in open-loop, the offset of the multisine signal plays an important role in determining the operating range of interest. It was selected heuristically such that the throttle operates in its normal operating range between 10° and 90° for the maximum possible duration and it does not stay too long at its hard mechanical limits. Having excited the throttle with the multisine signal, the measurement of the output signal is recorded and is displayed in Fig. 2.4. Next, we briefly describe the design procedure for this multisine signal. For further details, see [129].

A multisine signal $u(t)$ with a limited frequency bandwidth is determined by a finite number of N_h harmonics and is mathematically represented by the following equation:

$$u(t) = \sum_{i=1}^{N_h} a_{h,i} \cos(\omega_{h,i} \cdot t + \phi_{h,i}) + u_{\text{offset}}, \tag{2.3}$$

where u_{offset}, $a_{h,i}$, $\omega_{h,i}$ and $\phi_{h,i}$ denote the offset, amplitude, angular frequency and phase shift of the i-th harmonic component, respectively. When selecting the length of the input signal, one has to make a comprise between the model estimation quality and the computational tractability. On one hand, if the signal length is too small, the model may not be sufficiently excited (the problem of under excitation may occur) and will thus have poor generalization capability in regions where the system is not adequately excited. On the other hand, if the signal length is too large, data processing requires a long time, and thus the computational complexity is increased. Keeping these facts in mind, the length of the multisine signal is chosen to be 100 s. The frequency band of interest lies between $F_{\max} \approx 5$ Hz $\leq F \leq F_{\min} \approx 0.7$ Hz [95, 97]. In this frequency band, only the prime harmonics are considered so that the effects of nonlinear distortions can be minimized [93, 95, 97]. With regard to the amplitude ($a_{h,i} = 8,\ \forall i$) and the offset ($u_{\text{offset}} = 14$), they were selected heuristically such that they excite all the important amplitude and frequency characteristics of the system while staying with the operating range for the maximum duration and avoiding staying at the mechanical hard limits of opening and closing too often. In order to enhance the information content of the input-output data for the given fixed length of the input signal, the peak factor of the input signal is optimized by optimizing the phase values. Furthermore, one of the motivation behind optimizing the peak factor of the input multisine signal is that it increases the SNR (Signal to Noise ratio) of the input, thus allowing the maximum power injecting into the system for the given operating range of the input signal. For that matter, the Schröder phases [102] are used as the initial value for the optimization. The Schröder phases for a multisine signal having d spectral components is mathematically described by the following equation:

$$\begin{aligned} \phi_{h,i} &= \phi_{h,1} - i \cdot (i-1) \cdot \pi/d, \\ \phi_{h,1} &= 0, \\ 2 \leq i &\leq d \end{aligned} \tag{2.4}$$

In order to get a uniform coverage of the input signal in its operating region, the phase values of the multisine signal are optimized by iteratively minimizing the Covering Index (CovInd) [26, 116] as carried out by Ren et al. [97]:

$$\text{CovInd} = \frac{1}{N \cdot (1 - \frac{1}{n_z})} \sum_{c=1}^{n_z} (n_c - n_i)^2, \quad \text{with } n_i = \frac{N}{n_z} \tag{2.5}$$

where $N = 10000$ is the total number of datapoints, $n_z = 100$ is the number of intervals, $n_c = 100$ and n_i are the current and ideal number of data points inside each interval.

A skyline signal, shown in Fig. 2.5, is designed for the model validation. This test signal is typical in the industrial applications of throttles. It tests the modeling performance of the throttle in typical operation including the hard mechanical limits. This signal consists of random jumps of different amplitudes and durations. The amplitude of the input signal is chosen randomly in the range of 0 % to 100 %. This results in the output signal shown in Fig. 2.5.

2.3 Servo-Pneumatic Longitudinal Drive

2.3.1 Motivation

The second experimental case study uses a piston rod-less servo-pneumatic longitudinal drive which is also available in the laboratory of the Department of Measurement and Control Engineering, University of Kassel. Pneumatic drives have been used extensively in industries, thanks to advantages such as readily available pressured air as supply energy in manufacturing plants, low cost, and safety in areas with explosive atmospheres. The rationale for choosing this experimental setup for demonstrating the developed modeling approach is that the open-loop output of this system shows significant variability when the same input signal is applied to the system under similar initial conditions. The variability is particularly predominant at low frequencies, whereas at very high frequencies, the effect was hardly seen. These uncertainties may be attributed to complex stochastic frictional effects [96, 133].

With respect to classical modeling, Bernd et al. [9, 10] developed LS-optimal T1 fuzzy model and Multi-layer Perceptron (MLP) Artificial Neural Network (ANN) from crisp or deterministic input-output data sets to model the servo-pneumatic longitudinal drive. The modeling methods described in this dissertation in the subsequent chapters extend this

classical modeling case using crisp data to the novel interval modeling case using interval data in order to handle uncertainties due to the variability in the output.

2.3.2 Description of the Test Stand

The chosen test stand of the piston rod-less servo-pneumatic longitudinal drive of the company Festo Didactic (Festo-didactic-PneuPos 1/192490) is shown in Fig.2.6. In order to

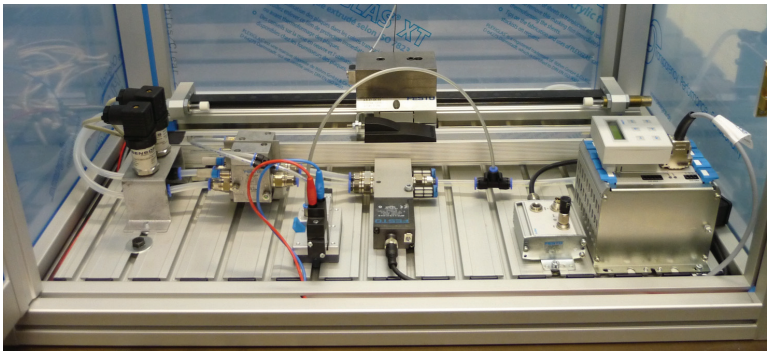


Figure 2.6: Test stand with piston rod-less linear servo-pneumatic drive

simplify the system, only the Single-Input-Single-Output (SISO) case of the system is considered. The input signal of the drive is taken as the voltage signal applied to the servo valves, whereas the output signal is taken as the longitudinal position of the piston rod. The

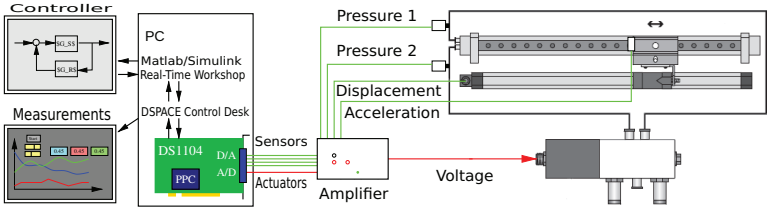


Figure 2.7: The experimental setup of the piston rod-less servo-pneumatic longitudinal drive

complete experimental setup of the test stand is shown in Fig. 2.7. As mentioned before, the drive does not have any pistons. Moreover, for forward and backward strokes, the drive consists of double acting cylinders. On account of the fact that the drive is piston-less and both the cylinder chambers are equal in size, the rod of the drive stands still, when equal force is exerted on the chambers, as a result of equal chamber pressure.

2.3.3 Design of Experiment

By using the same guidelines as in case of the electro-mechanical throttle valve, the sampling time for the servo-pneumatic longitudinal drive is chosen as $T_s = 100$ ms (using $T_s = 10$ ms does not provide for any noticeable improvement of the model). As in case of the electro-mechanical throttle, a phase optimized band-limited multisine signal with the maximum frequency of $f_{\max} = 1$ Hz is chosen as the input signal. The first half of the data is used for identification and the later half for validation. The total duration of the signal is chosen to be $T = 100$ s, which leads to $N = T/T_s = 100/0.1 = 1000$ samples. The designed multisine input signal and the corresponding output signal is shown in Fig. 2.8.

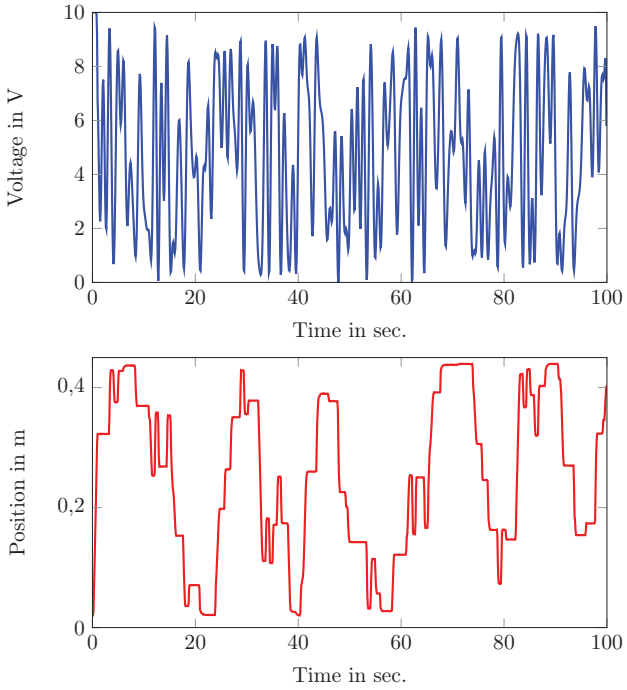


Figure 2.8: Multisine input signal and the corresponding output signal

2.4 Summary and Discussion

In this chapter, a brief description of the academic benchmark problem and the two experimental test stands, along with the motivation and the details of the design of experiment, have been given and the corresponding references have been provided.

3

Uncertainty Analysis of Nonlinear Discrete Dynamic Systems

In this chapter, some techniques to analyse uncertainties are presented. Specifically, stochastic uncertainties are modelled in the realm of statistics and probability theory. The techniques of normalized histogram and Kernel Density Estimation (KDE) are used for the non-parametric estimation, and Gaussian Mixture Models (GMMs) [69] for the semi-parametric description of uncertainty. These methods are briefly described in this chapter. Since many modelling methods make the normality assumption regarding the uncertainty, some normality tests will be presented. These tests describe how well the underlying distribution resembles the normal or Gaussian distribution. As discussed in Chapter 1, the developed modelling methods assume either crisp or interval data. By using the statistical and probability techniques presented in this chapter, the output time series is converted to interval form. This interval response is described by envelopes in which the actual response is expected to lie with a certain confidence level. Some techniques of developing those envelopes are discussed next. These techniques include the worst case (or max-min) envelopes, percentile based envelopes, and confidence interval based envelopes. Finally, these techniques are applied in the experimental and simulation case studies. The chapter ends with a brief discussion.

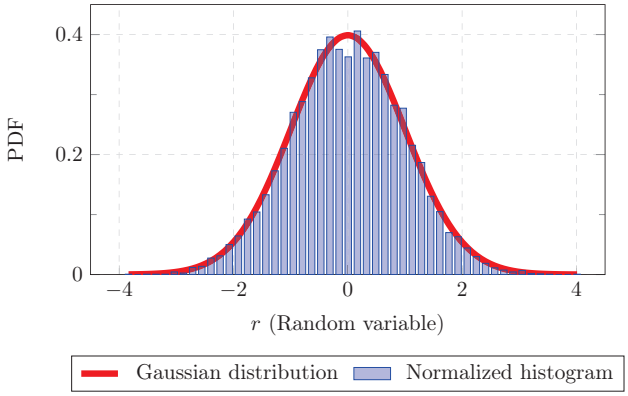


Figure 3.1: Illustration of a normalized histogram as an estimate of the normal distribution

3.1 Probability Density Estimation

Probability density function (PDF) estimation deals with constructing an estimate of an unobservable underlying PDF based on observed data, where the data is considered to be a random sample originating from that Probability Density Function (PDF). The non-parametric (normalized histogram, kernel density estimation) and semi-parametric (GMMs) used in this research are described in the sequel.

3.1.1 Normalized Histogram

This technique is considered as the oldest and most widely used technique for estimating a PDF. Consider a continuous random variable r defined on some probability space $(\Omega_\mathsf{r}, \mathcal{B}_\mathsf{r}, P_\mathsf{r})$, where Ω_r is the sample space, $\mathcal{B}_\mathsf{r}$ is the Borel sigma-algebra, and P_r is the probability measure of r. Let r denote the scalar value of the random variable r. Denoting $p_\mathsf{r}(r)$ as the PDF of the random variable r, the probability P_r that r lies between a and b, where $a, b \in \mathbb{R}$ and $a < b$, is given by:

$$P_\mathsf{r}(a < \mathsf{r} < b) = \int_{r=a}^{b} p_\mathsf{r}(r)dr. \tag{3.1}$$

Assuming r_0 and h_b denote the origin and bin width of a histogram, respectively, the bins of the histogram are defined as intervals of the form $[r_0 + n_I h_b, r_0 + (n_I + 1)h_b]$, where n_I belongs to the set of integers $\mathbb{N}$. Let $(r_1, \ldots, r_{N_o})$ be an independent and identically

distributed (i.i.d.) sample of the random variable r drawn from $p_{\mathsf{r}}(r)$, where N_o is the total number of observations. The normalized histogram $\hat{p}_{\mathsf{r}}^{\mathrm{NH}}(r)$ which can be seen as a rough and discrete approximate of $p_{\mathsf{r}}(r)$ is defined by

$$\hat{p}_{\mathsf{r}}^{\mathrm{NH}}(r) = \frac{1}{h_b \cdot N_{\mathrm{o}}} \sum_{i=1}^{N_o} I_{[r_0+n_I h_b, r_0+(n_I+1)h_b]}(r = r_i), \tag{3.2}$$

where $I(\cdot)$ is the indicator function defined as follows:

$$I_{[r_0+n_I h_b, r_0+(n_I+1)h_b]}(r = r_i) = \begin{cases} 1 & \textbf{if } r_i \in [r_0 + n_I h_b, r_0 + (n_I+1)h_b] \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

When $h_b \to 0$ & $N_o \to \infty$, $\hat{p}_{\mathsf{r}}^{\mathrm{NH}}(r) \to p_{\mathsf{r}}(r)$. For a PDF, the total area under the curve integrates to 1. Note that, by simply dividing the frequency histogram by the total number of observations (N_o), one gets a relative frequency histogram. The bin heights add up to unity but the area under all bins will not match unity (unless $h_b = 1$). To get a rough discrete estimate of the PDF, the relative frequency histogram must further be divided by the value of bin width h_b.

Example 3.1.1. *Fig. 3.1 shows an example of a normalized histogram. As it is apparent from the figure, the normalized histogram reasonably approximates the Gaussian PDF with zero mean and unit variance, provided that the bin width is chosen appropriately.*

3.1.2 Kernel Density Estimation

The PDF estimation using a normalized histogram has the problem that $\hat{p}_r^{\mathrm{NH}}(r)$ is not a smooth function of r. The problem can be avoided by using another well-known non-parametric technique known as Kernel Density Estimation (KDE). The kernel density estimator with kernel Ker is defined by:

$$\hat{p}_{\mathsf{r}}^{\mathrm{KDE}}(r) = \frac{1}{h_w \cdot N_{\mathrm{o}}} \sum_{i=1}^{N_o} Ker\left(\frac{r - r_i}{h_w}\right), \tag{3.4}$$

where h_w is the window width (also known as smoothing parameter or bandwidth), and Ker is usually a symmetric function, such as the Gaussian function, which satisfies the following condition:

$$\int_{-\infty}^{\infty} Ker(r)dr = 1. \tag{3.5}$$

Kernel	$Ker(w)$
Uniform	$\frac{1}{2}I(\|w\| \leq 1)$
Triangular	$(1-\|w\|)I(\|w\| \leq 1)$
Epanechnikov	$\frac{4}{3}(1-w^2)I(\|w\| \leq 1)$
Quartic	$\frac{15}{16}(1-w^2)^2I(\|w\| \leq 1)$
Triweight	$\frac{35}{32}(1-w^2)^3I(\|w\| \leq 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}w^2)$
Cosine	$\frac{\pi}{4}\cos(\frac{\pi}{2}w)I(\|w\| \leq 1)$

Table 3.1.: Various kernels for KDE

Intuitively, the kernel density estimator can be considered as a sum of 'bumps' placed at the observations, where the shape of the bumps is determined by Ker, while h_w determines their width [104]. The value of h_w is selected based on the compromise between under smoothing and over smoothing On one hand, the value of h_w that is too small causes too many spurious artificats, while on the other hand the value of h_w that is too big obscures much of the underlying information. A comprehensive review of fully automatic selectors of bandwidth h_w for KDE is given in [31]. Even though Gaussian kernels are most often used, there are various kernels as shown in Tab. 3.1. Here $w = \frac{r-r_i}{h_w}$, and $I(|w| \leq 1)$ represent an indicator function, which is mathematically defined as follows:

$$I(|w| \leq 1) = \begin{cases} 1 & \text{if } |w| \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

Histograms and estimates based on KDE are closely related, but the latter is equipped with some nice properties such as smoothness or continuity on account of using a suitable kernel.

Example 3.1.2. *To compare the relation between a histogram and a kernel density estimate, we consider an example of PDF estimation using* $N_o = 6$ *data points:* $r_1 = -3$, $r_2 = -1$, $r_3 = -0.5$, $r_4 = 1.8$, $r_5 = 5$, *and* $r_6 = 7$. *For this purpose, we select the bin width* $h_b = 2$ *and origin* $r_0 = -4$, *for the given data set. Thus, we get a total of 6 bins* ($n_I = 0, 1, \ldots, 5$), *each having the width of 2 as shown in Tab. 3.2. We normalize the frequency histogram by dividing it with the factor* $h_b \cdot N_o = 2 \cdot 6 = 12$ *in accordance with* (3.2). *For the KDE, we choose a Gaussian kernel* $Ker_{\text{GAUSS}}(w) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}w^2)$, *where* $w = \frac{r-r_i}{h_w}$, *with* h_w *is the kernel width*

Integer: n_I	Interval: $[r_0 + n_I h_b, r_0 + (n_I + 1)h_b]$	No. of r_i in the interval
0	[-4,-2]	1
1	[-2,0]	2
2	[0,2]	1
3	[2,4]	0
4	[4,6]	1
5	[6,8]	1

Table 3.2.: Intervals (bins) for the normalized histogram and the corresponding frequencies

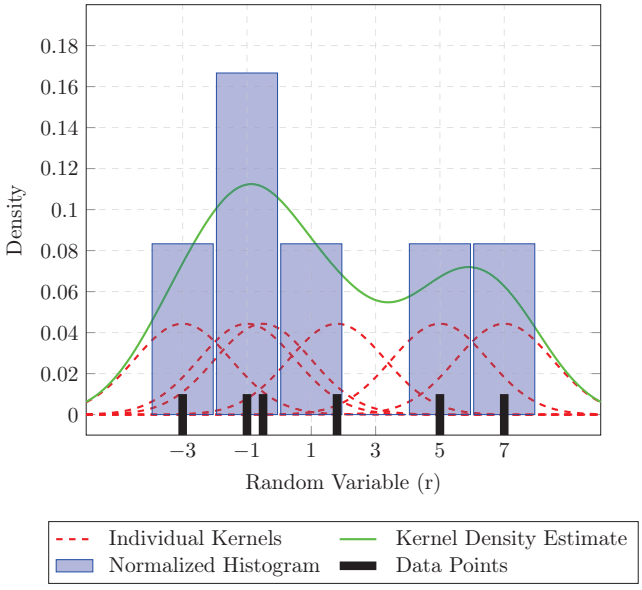


Figure 3.2: Normalized histogram and kernel density estimate

σ, whose value is chosen as $h_w = \sigma = 1.5$ and r_i is the i-th data point, $i = 1, \ldots, N_o$. The individual kernels and their sum are illustrated in Fig. 3.2.

In this work, the considered kernel is Gaussian, which uses a window parameter (or width) that is a function of the number of points in r. Moreover, the PDF is evaluated at 100

equally spaced points that cover the entire range of the data. This approach works best with continuously distributed samples. Considering the case of approximating univariate data with Gaussian kernels and assuming that the underlying PDF being approximated is Gaussian, the optimal choice for h_w (i.e. the bandwidth that minimizes the mean integrated squared error) [104] is given by

$$h_w = \left(\frac{4\sigma^5}{3N_o}\right)^{\frac{1}{5}} \approx 1.06\hat{\sigma}N_o^{-\frac{1}{5}} \tag{3.7}$$

where $\hat{\sigma}$ denotes the standard deviation of the sample. This approximation is known as the normal distribution/Gaussian approximation or Silverman's [104] rule of thumb.

3.1.3 Gaussian Mixture Models

Gaussian Mixture Models (GMMs) are among the most statistically matured methods that have been widely used for semi-parametric PDF estimation and clustering. In many practical situations, the observed PDF often cannot be described to a fair extent by a single Gaussian distribution. In such cases, it is advantageous to describe it in terms of a GMM.

A GMM is basically a weighted combination of two or more Gaussian components (also called mixtures). Mostly, these models are estimated by the well known Expectation Maximization (EM) algorithm in the framework of Maximum Likelihood Estimation (MLE). Theoretically, any continuous PDF can be described completely by a GMM of infinite mixtures. However for practical purposes, only finite mixtures of GMMs are considered that reasonably approximate the underlying PDF.

Consider a vector $\mathbf{r} \in \mathbb{R}^d$ originated from a super population Q, which is a mixture of a finite number of Gaussian sub-populations $Q_1, ..., Q_q$ in some mixing proportions $\pi_1, ..., \pi_q$, respectively, where

$$\sum_{l=1}^{q} \pi_l = 1 \text{ and } 0 \leq \pi_l \leq 1.$$

The PDF of $\mathbf{r}$ is given by

$$\begin{aligned} p(\mathbf{r}; \boldsymbol{\theta}_{\text{GMM}}) &= \sum_{l=1}^{q} \pi_l \cdot p(\mathbf{r}|l; \boldsymbol{\phi}_{\text{GMM}_l}) \\ &= \sum_{l=1}^{q} \pi_l \cdot \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_l|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_l)^T(\Sigma_l)^{-1}(\mathbf{r}-\boldsymbol{\mu}_l)\right), \end{aligned} \tag{3.8}$$

where π_l is the mixing coefficient, $p(\mathbf{r}; \boldsymbol{\phi}_{\text{GMM}_l})$ is the PDF corresponding to Q_l and $\boldsymbol{\theta}_{\text{GMM}}$ represents the vector of all unknown parameters for these q component densities. For the case of Gaussian functions as component densities, $\boldsymbol{\phi}_{\text{GMM}_l}$ contains the mean vectors $\boldsymbol{\mu}_l$, and covariance matrices Σ_l for $l = 1, ..., q$. The vector $\boldsymbol{\theta}_{\text{GMM}}$ lumping all unknown parameters belongs to some parameter space Θ_{GMM} and is estimated using the EM algorithm [22, 121].

3.1.3.1 Expectation Maximization Algorithm

The Expectation Maximization (EM) algorithm [22] is one of the most widely used algorithms for unsupervised learning. It is used to iteratively find the Maximum Likelihood (ML) or Maximum A Posterior (MAP) estimates of parameters of a probabilistic or statistical model, wherein the model has some unobserved latent or missing variables. As the name suggests, the EM algorithm iterates between two steps, namely Expectation (E) and Maximization (M) steps. In the E step, an expectation function of the log-likelihood is created based on the current estimates of the parameters. By maximizing this expectation function, a new estimate of parameters is calculated in the M step. The process repeats until a local maximum is reached.

Let $R = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_{N_o})$ be a set of i.i.d. observed data, $S = (s_1, s_2, \ldots, s_{N_o})$ be a set of corresponding unobserved latent or missing data, and $\boldsymbol{\theta}$ be a vector of unknown parameters. Further, assume that the likelihood function is represented as $\mathcal{L}(\boldsymbol{\theta}; R, S) = p(R, S|\boldsymbol{\theta})$, the MLE of $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta}_{\text{MLE}}$, is given by the marginal likelihood of R

$$\mathcal{L}(\boldsymbol{\theta}_{\text{MLE}}; R) = p(R|\boldsymbol{\theta}_{\text{MLE}}) = \sum_S p(R, S|\boldsymbol{\theta}_{\text{MLE}}) \tag{3.9}$$

The EM algorithm aims to find the MLE of the marginal likelihood given in (3.9) by iteratively applying the E and M steps as follows:

1. E (Expectation) step: In this step, given the observed data R and the current estimate $\boldsymbol{\theta}^{(k)}_{\text{MLE}}$, the expected value of the log likelihood, with respect to the conditional distribution of S given R and $\boldsymbol{\theta}^{(k)}_{\text{MLE}}$, is calculated as follows:

$$Q(\boldsymbol{\theta}_{\text{MLE}}|\boldsymbol{\theta}^{(k)}_{\text{MLE}}) = \text{E}_{R|S,\boldsymbol{\theta}^{(k)}_{\text{MLE}}} \left[\log \mathcal{L}(\boldsymbol{\theta}_{\text{MLE}}; R, S)\right] \tag{3.10}$$

2. M (Maximization) step: Assuming the missing data S is known, the likelihood function is maximized in this step. The updated parameter vector $\boldsymbol{\theta}_{\text{MLE}}^{(k+1)}$ that maximizes this likelihood function is given by

$$\boldsymbol{\theta}_{\text{MLE}}^{(k+1)} = \arg\max_{\boldsymbol{\theta}_{\text{MLE}}} Q(\boldsymbol{\theta}_{\text{MLE}}|\boldsymbol{\theta}_{\text{MLE}}^{(k)}) \tag{3.11}$$

In the special case of GMM, a mixture model consisting of two mixtures or components ($q = 2$) is considered to illustrate the application of EM in finding the parameters of a GMM. The observed data $\boldsymbol{r}_i$, $i = 1, \ldots, N_o$ can originate from either of the two Gaussian components $\mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1)$ and $\mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2)$. The corresponding value of s_i determines the component from which $\boldsymbol{r}_i$ originates.

$$\boldsymbol{r}_i|(s_i = 1) \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1) \quad \text{and} \quad \boldsymbol{r}_i|(s_i = 2) \sim \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2) \tag{3.12}$$

The probability that s_i takes the value 1 or 2 is determined by the mixing coefficients π_i, i.e.

$$P(s_i = 1) = \pi_1 \quad \text{and} \quad P(s_i = 2) = \pi_2 = 1 - \pi_1 \tag{3.13}$$

The task is to estimate $\boldsymbol{\theta}_{\text{GMM}} = (\pi_1, \pi_2, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2)$ using the EM algorithm, and thus in this case $\boldsymbol{\theta}_{\text{GMM}} = \boldsymbol{\theta}_{\text{MLE}}$. For that purpose, the likelihood function is calculated:

$$\mathcal{L}(\boldsymbol{\theta}_{\text{GMM}}; R, S) = p(R, S|\boldsymbol{\theta}) = \prod_{i=1}^{N_o} \sum_{l=1}^{2} I(s_i = l)\pi_l p(\boldsymbol{r}_i; \boldsymbol{\mu}_l, \Sigma_l) \tag{3.14}$$

where $I(\cdot)$ is an indicator function and $p(\cdot)$ is the PDF of a multivariate Gaussian distribution. Next, an error function is defined as the negative log likelihood as follows:

$$E = -\log \mathcal{L} \tag{3.15}$$

Accordingly, minimizing E is equivalent to maximizing $\mathcal{L}$. By setting the derivatives of the change in error function $(E^{(k+1)} - E^{(k)})$ to zero (see [92] for details), the following update equations are obtained for the parameters of the GMM:

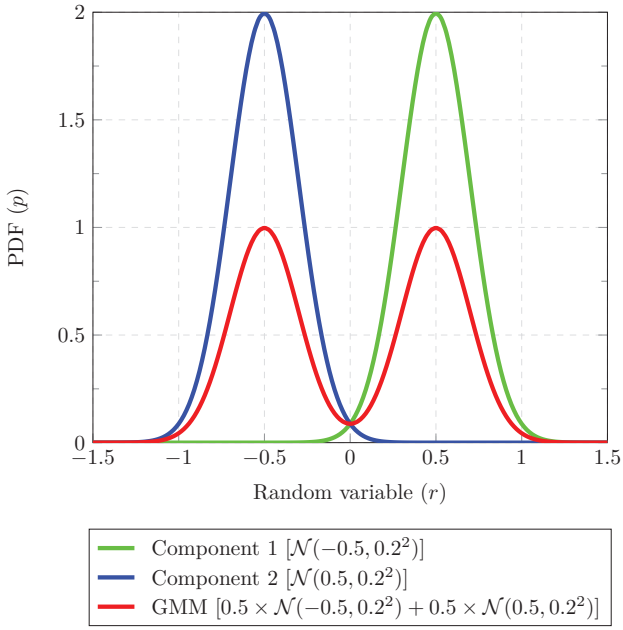


Figure 3.3: PDF estimation by a GMM and the corresponding mixtures

$$\mu_l^{(t+1)} = \frac{\sum_{i=1}^{N_o} p^{(t)}(l|\boldsymbol{r}_i)\boldsymbol{r}_i}{\sum_{i=1}^{N_o} p^{(t)}(l|\boldsymbol{r}_i)} \tag{3.16}$$

$$\Sigma_l^{(t+1)} = \frac{\sum_{i=1}^{N_o} p^{(t)}(l|\boldsymbol{r}_i)\left[\boldsymbol{r}_i - \boldsymbol{\mu}_l^{(t)}\right]\left[\boldsymbol{r}_i - \boldsymbol{\mu}_l^{(t)}\right]^\top}{\sum_{i=1}^{N_o} p^{(t)}(l|\boldsymbol{r}_i)} \tag{3.17}$$

$$\pi_l^{(t+1)} = \frac{1}{N_o}\sum_{i=1}^{N_o} p^{(t)}(l|\boldsymbol{r}_i) \tag{3.18}$$

$$\text{where,} \quad p^{(t)}(l|\boldsymbol{r}_i) = \frac{p^{(t)}(\boldsymbol{r}_i|l)\pi_l^{(t)}}{p^{(t)}(\boldsymbol{r}_i)} \tag{3.19}$$

Example 3.1.3. *Fig. 3.3 illustrates a simple example of a PDF estimation by GMM using two mixtures. Both the mixing components are set to 0.5, so half of the contribution comes from each mixture. This is an example of a bimodal distribution, since it requires two*

mixtures to estimate the underlying PDF. The parameters of the Gaussian mixtures as well as the GMM are shown in Fig. 3.3.

3.1.3.2 Model Selection Methods

In order to determine the number of Gaussian components or mixtures required to adequately represent the true PDF of the underlying data, various model selection methods can be used. Not only do these model selection techniques have the capability of providing a mechanism to minimize the log-likelihood function of the incomplete data, but they also avoid the undesirable effect of overfitting. Among various techniques existing in the literature, including the exhaustive n-fold cross-validation [6] technique, two well-known techniques namely the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC) will be given next.

The AIC technique [4] tends to simultaneously minimize the negative log-likelihood function of the incomplete data and the total number of parameters used in the model. It specifically favours parsimonious models having both low number of parameters and a reasonably low value of negative log-likelihood function for a given model class. It is defined as:

$$AIC(\boldsymbol{\theta}) = -2 \cdot \log p(R|\boldsymbol{\theta}) + 2 \cdot N_p \tag{3.20}$$

where $\boldsymbol{\theta}$ denotes the parameter vector of the model, R is the input data, and N_p is the total number of parameters in the model. The model having the lowest value of AIC is considered as the best one. However, in cases where the number of input vectors, N_o, is small relative to N_p, AIC has the tendency to provide a negatively-biased estimate of the difference between the true distribution and the model. To take that into consideration, a corrected AIC (AICc) has been proposed in [33], which is given by

$$AIC_c(\boldsymbol{\theta}) = -2 \cdot \log p(R|\boldsymbol{\theta}) + 2 \cdot N_p \left(\frac{N_o}{N_o - N_p - 1} \right) \tag{3.21}$$

As previously, the model that minimizes the AIC_c will be selected. AIC_c is particular useful when the number of observations N_o is limited and it tends to outperform AIC in such cases.

Finally, the BIC is defined mathematically as:

$$BIC(\boldsymbol{\theta}) = -2 \cdot \log(R|\boldsymbol{\theta}) + N_p \cdot \log(N_o) \tag{3.22}$$

Again the model that minimizes the BIC will be chosen.

3.2 Testing Data for Normality

Most statistical procedures use assumptions about the normality of data. For instance, the distribution of the residuals in linear regression analysis is usually assumed to be normal. When the underlying assumption of the normality of data is not met, the interpretation and inferences that are based on the normality assumption may not be reliable or valid. Hence, it is important to check for this assumption before proceeding with any relevant statistical procedure. Three methods are commonly used in literature for carrying out normality tests. The most common and the easiest test consists of graphical methods which include histogram, box-plot, normal quantile-quantile plot (Q-Q plot), stem-and-leaf plot etc. Although graphical methods can be used as a useful tool to visually inspect the data for normality, they are still not sufficient to provide conclusive evidence that the normality assumption holds. Consequently, to support the graphical methods, more formal methods which are numerical methods and normality tests should be performed before making any conclusion about the normality of the data. The numerical methods include the skewness and kurtosis coefficient whereas a normality test is a more formal procedure which involves testing whether a particular data follows a normal distribution. There is a significant amount of normality tests available in the literature. However, the most common normality test procedures are the Shapiro-Wilk (SW) test [103], Kolmogorov-Smirnov (KS) test [17], Anderson-Darling (AD) test [5] and Lillifors (LF) test [63].

In this work, two methods have been employed for checking normality of data which are the normality probability plot and the SW test. The purpose of a normal probability plot is to graphically assess whether the underlying data can be described by a normal distribution. If the data are normal, the plot will be linear. Other distribution types will introduce curvature in the plot. As normality test, the SW test has been used; since it has proven to be the most powerful normality test for a given significance when compared with KS, LF, and AD tests via Monte Carlo simulation of sample data originated from symmetric and asymmetric distributions [91].

Given a sample $r_1, \ldots, r_{N_o}$ of a random variable r of N_o real-valued observations, the SW test is a test of the composite hypothesis that the data are i.i.d. and normal, i.e. $\mathcal{N}(\mu, \sigma^2)$, for

some unknown real μ and some $\sigma > 0$. Given an ordered random sample, $r_1 < r_2 < \ldots < r_{N_o}$, the original SW test statistic is defined as [103]:

$$W = \frac{(\sum_{i=1}^{N_o} a_i^{\mathrm{SW}} \cdot r_i)^2}{\sum_{i=1}^{N_o}(r_i - \bar{r})^2} \tag{3.23}$$

where r_i is the i-th order sample, $\bar{r}$ is the sample mean, and a_i^{SW} is the i-th component of the constant vector $\boldsymbol{a}^{\mathrm{SW}} = [a_1^{\mathrm{SW}}, \ldots, a_{N_o}^{\mathrm{SW}}]^\top$, which is given by:

$$\boldsymbol{a}^{\mathrm{SW}} = \frac{\boldsymbol{m}^\top \Sigma^{-1}}{(\boldsymbol{m}^\top \Sigma^{-1} \Sigma^{-1} \boldsymbol{m})^{\frac{1}{2}}} \tag{3.24}$$

where $\boldsymbol{m} = [m_1, \ldots, m_n]^\top$, and $m_1, \ldots, m_{N_o}$ are the expected values of the order statistics of i.i.d r sampled from the standard normal distribution, and Σ is the covariance matrix of those order statistics.

Like every hypothesis test, the application of SW test consists of a test statistic W and a p-value. If the p-value is less than, say, the conventional level of 0.05, one rejects the normality hypothesis, otherwise one does not reject it. The value of W always satisfies $0 < W < 1$. For values of W close enough to unity (depending on N_o), the normality hypothesis will not be rejected. For smaller W, it will be rejected.

Example 3.2.1. [1] *To illustrate the normality tests using normality probability plots (Statistics and Machine Learning Matlab Toolbox: normplot) and Shapiro Wilk test, 1000 random numbers have been generated from four different distributions, namely a standard normal distribution ($\mu = 0$, $\sigma = 1$), a fat-tailed distribution (Student's t distribution with five degrees of freedom), a right-skewed distribution (Pearson random numbers with $\mu = 0$, $\sigma = 1$, skewness=0.5, kurtosis=3), and a left-skewed distribution (Pearson random numbers with $\mu = 0$, $\sigma = 1$, skewness=-0.5, kurtosis=3). Fig. 3.4 shows the histograms of these distributions along with the results of SW tests.*

For the SW test, $H = 0$ (or $H = 1$) implies that the null hypothesis is not rejected (or rejected) at the significance level α ($\alpha = 0.05$ is chosen in this example). Higher (or lower) values of p signify how much the given data is likely (or unlikely) with a true null hypothesis. The normal probability plots in Fig. 3.5 visually demonstrate how likely the given distributions

[1] The example has been taken and modified from the Matlab documentation page (`http://de.mathworks.com/help/stats/normplot.html`) and for the SW test the code has been taken from MathWorks file Exchange (File ID#13964, `https://www.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests`).

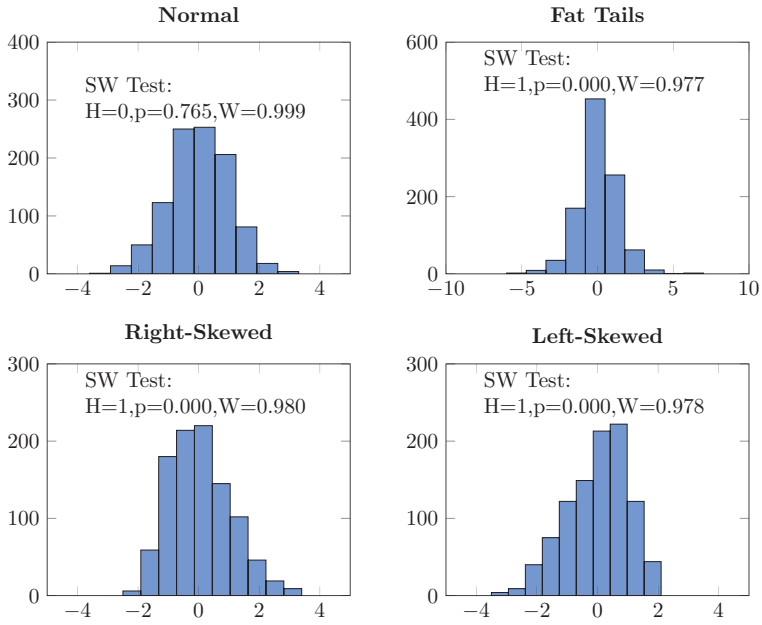


Figure 3.4: Histograms of the chosen distributions and the result of SW test

are to be normal. The figure clearly shows that only in the first case, the given data points are lying on the straight line, and thus the normal distribution is correctly identified.

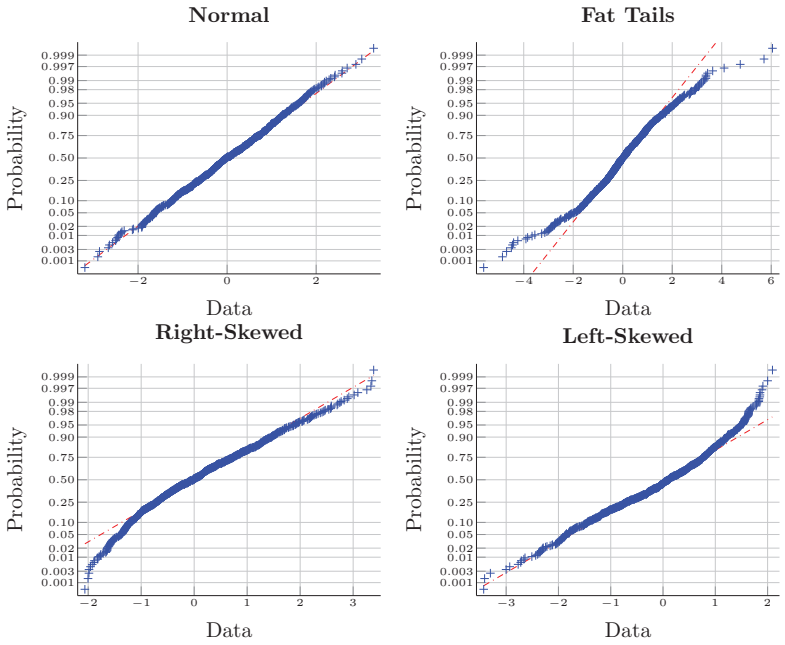


Figure 3.5: Normal probability plots for the given distributions

3.3 Stochastic Treatment of Stochastic Dynamic Systems

To keep things and notations simple, we consider a discrete time Single-Input Single-Output (SISO) dynamic system. The theory can easily be extended to the case of Multiple-Input Single-Output (MISO). Furthermore, a Multiple-Input-Multiple-Output (MIMO) system can be considered as a combination of several MISO systems. As discussed before, this research is concerned with the case of deterministic inputs on account of the fact that the input signal is exactly reproducible in each identification experiment. The underlying system can be mathematically defined as follows [131]:

$$y_k = \mathcal{F}(x_k) + \zeta_k, \quad k = 1, \ldots, N \tag{3.25}$$

where,

- N and k denote the length of experiment (total number of observations) and the discrete time index, respectively.

- y_k is a scalar-valued stochastic output signal

- x_k is a vector-valued stochastic regressor quantity, which consists of tapped delay lines of output (stochastic) and input (deterministic) signals, i.e., $\mathsf{x}_k = [\mathsf{y}_{k-1}, \ldots, \mathsf{y}_{k-n_y}, u_{k-\tau-1}, \ldots, u_{k-\tau-n_u}]^\top$, where τ, n_u and n_y denote the dead time, the number of lagged input samples, and the number of lagged output samples, respectively.

- $\mathcal{F}(\cdot)$ is a stochastic process governing the underlying stochastic dynamics

- ζ_k is an additive noise term with zero mean ($\mu_{\zeta_k} = 0$) and finite variance ($\sigma_{\zeta_k} < \infty$)

In practice, $\mathcal{F}(\mathsf{x}_k)$ and ζ_k come from two independent processes; thus, they are independent of each other. Given a probability space $(\Omega_{\mathsf{y}_k}, \mathcal{B}_{\mathsf{y}_k}, \Pr_{\mathsf{y}_k})$; where Ω_{y_k} is the sample space, $\mathcal{B}_{\mathsf{y}_k}$ is the Borel sigma-algebra, and $\Pr_{\mathsf{y}_k}$ is a probability measure; the underlying system can be seen as a discrete stochastic process of the form

$$\{\mathsf{y}_k | \mathsf{y}_{k-1}, \ldots, \mathsf{y}_{k-n_y}, u_{k-\tau-1}, \ldots, u_{k-\tau-n_u} : k \in K\}, \tag{3.26}$$

where y_k is a sequence of random variables, i.e. $\mathsf{y}_k = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_N\}$, conditioned on the tapped delay lines of inputs and outputs, and $K = \{1, \ldots, N\}$ is called the index set of y_k. We give a formal definition of a dynamic stochastic process next.

Definition 3.3.1. *(Discrete dynamic stochastic process): A discrete dynamic stochastic process is a sequence of discrete random variables* $\mathsf{y}(k)$*'s conditioned on the tapped delay lines of inputs and outputs defined on some probability space* $(\Omega_{\mathsf{y}_k}, \mathcal{B}_{\mathsf{y}_k}, \Pr_{\mathsf{y}_k})$ *and ordered with respect to a time index* k*, where* k *takes values in an index set* K*.*

Thanks to the time-homogeneity [90] assumption of the underlying stochastic process, every observation $y(k)$ in the time series is drawn from the same random variable $\mathsf{y}(k)$. This assumption is reasonable for time invariant systems where the data generating mechanism has a constant PDF of $\mathsf{y}(k)$ over time. This property is called strict stationarity of the stochastic process.

Definition 3.3.2. *(Strict stationarity): The dynamic stochastic process*

$$\{\mathsf{y}_k | \mathsf{y}_{k-1}, \ldots, \mathsf{y}_{k-n_y}, u_{k-\tau-1}, \ldots, u_{k-\tau-n_u}\}_{k \in K}$$

is strictly stationary if for all $N \in \mathbb{N}$ *and* $h \in \mathbb{Z}$

$$\{\mathsf{y}_k|\mathsf{y}_{k-1},\ldots,\mathsf{y}_{k-n_y},u_{k-\tau-1},\ldots,u_{k-\tau-n_u},k\in K\} \overset{d}{=} \{\mathsf{y}_{k+h}|\mathsf{y}_{k+h-1},\ldots,\mathsf{y}_{k+h-n_y},u_{k-\tau+h-1},\ldots,u_{k-\tau+h-n_u},k\in K\} \quad (3.27)$$

where $\overset{d}{=}$ *denotes equality in distribution.*

The time invariance of only the first two moments (i.e. mean and variance) is called weak stationarity.

Definition 3.3.3. *(Weak stationarity): The dynamic stochastic process*

$$\{\mathsf{y}_k|\mathsf{y}_{k-1},\ldots,\mathsf{y}_{k-n_y},u_{k-\tau-1},\ldots,u_{k-\tau-n_u},k\in K\}$$

is weakly stationary if for all $h \in \mathbb{Z}$ *and* $k \in K$*,*

$$\mathbb{E}(\mathsf{y}_k) = \mu_k \quad (3.28)$$

$$\mathbb{C}\mathrm{ov}(\mathsf{y}_k,\mathsf{y}_{k-h}) = \gamma_k(h) \quad (3.29)$$

with $\gamma(0) < \infty$*, where* $\gamma_k(h)$ *is the autocovariance function* $\gamma_k(h) = \mathbb{E}[(\mathsf{y}_k - \mu_k)(\mathsf{y}_{k-h} - \mu_{k-h})]$*.*

Indeed, the strict stationarity automatically implies the weak stationarity provided that the variance is finite as described in the following theorem:

Theorem 3.3.1. *Let* $\{\mathsf{y}_k|\mathsf{y}_{k-1},\ldots,\mathsf{y}_{k-n_y},u(k-\tau-1),\ldots,u(k-\tau-n_u),k\in K\}$ *be a dynamic stochastic process. If this process is strictly stationary, then it is also weakly stationary if and only if* $\sigma_k^2 < \infty$*.*

Hence for the underlying stochastic dynamic system, the mean and variance of all elements are time invariant. (3.26) can be written in the form of N-tuple, i.e. $\mathsf{y}_k = (\mathsf{y}_1,\ldots,\mathsf{y}_N)$, which is a random variable taking values in $\mathbb{R}^N$. The distribution of y_k is a probability measure on $\mathbb{R}^N$. Since the experiment is repeated M times, the given stochastic process can be viewed as M realizations of y_k. Moreover, each y_k is assumed to be i.i.d., considering the fact that each experiment can be performed independently of each other. However, the components of the random vector $\mathbf{y}$ dependent upon each other through the tapped delay lines of inputs and outputs defined by the system dynamics as given by the conditional densities:

$$p(\mathsf{y}_k|\mathsf{y}_{k-1},\ldots,\mathsf{y}_{k-n_y},u_{k-\tau-1},\ldots,u_{k-\tau-n_u}) \quad (3.30)$$

For accurate treatment of uncertainty, these densities can be estimated for each instant k as described in detail in Sec. 3.1. To check the nature of the distribution followed by y_k, the

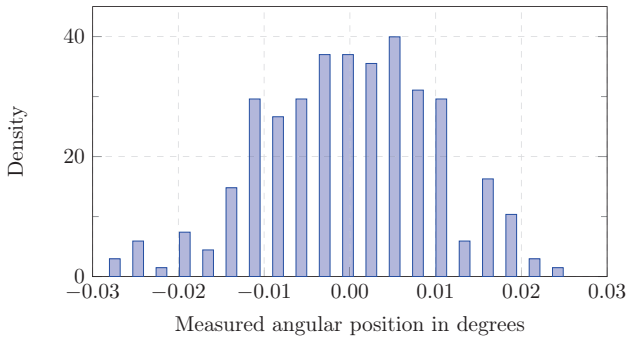


Figure 3.6: Normalized histogram of the noise signal for the electro-mechanical throttle case study

histogram of output values at each time instant can be plotted for visual inspection. As a particular case of normality, the distribution can be tested by normality tests described in Sec. 3.2. Having had the knowledge about the underlying distribution, a realistic coverage can be used for determining confidence interval based envelopes of the response, as it will be shown it in the next section.

The expected value ($E(\cdot)$) of y_k in (3.25) is given by

$$y^{\exp}(k) = E(\mathsf{y}_k) = E(\mathcal{F}(\mathsf{x}_k) + \zeta_k) = \mu_k \tag{3.31}$$

Using the variance sum law, the variance ($\sigma_{(\cdot)}$) of y_k in (3.25) can be calculated as:

$$\sigma^2_{\mathsf{y}_k} = \sigma^2_{\mathcal{F}(\mathsf{x}_k)} + \sigma^2_{\zeta_k} \tag{3.32}$$

where $\sigma^2_{\zeta_k}$, being the variance of noise, can be experimentally estimated by taking repeated measurements of the output signal while keeping the input signal constant. From the selected case studies it is evident that these systems have considerably lower measurement noise as compared to their inherent stochasticity. It is illustrated for the case of the electro-mechanical throttle in Fig. 3.6.

In such cases, the variance due to stochasticity is significantly larger than the variance of noise, i.e. $\sigma_{\zeta_k} \ll \sigma_{\mathcal{F}(\mathsf{x}_k)}$, which leads to $\sigma_{\mathsf{y}_k} \cong \sigma_{\mathcal{F}(\mathsf{x}_k)}$.

3.4 Determination of Envelopes of Response

As described in the previous chapter, one of the presented Type-1 (T1) and Interval Type-2 (IT2) models in this dissertation require interval data. In the proposed approach, the observed input signal is assumed to be deterministic because the input generation mechanism produces the same input signal each time in multiple experiments. Owing to stochastic factors such as friction, uncertainty in the output signal is observed. This uncertainty is primarily governed by the variability in the system output. By applying the techniques of statistics and the probability theory, the lower and upper bounds of response are obtained; we refer to them as envelopes of the response. These envelopes, together with the deterministic input, constitute interval data for identification. The methods to construct these envelopes are given in the sequel.

For a formal mathematical treatment, we consider an input-output data matrix $Z = [\boldsymbol{u}, \mathsf{Y}] \in \mathbb{R}^{N\times(M+1)}$ for model identification and validation, where $\boldsymbol{u} \in \mathbb{R}^N$ and $\mathsf{Y} \in \mathbb{R}^{N\times M}$ denote the input vector and the output matrix, respectively; N is the length of one experiment (i.e. the total number of data points in one experiment) and M is the total number of repeated experiments under the same initial condition and input. The input vector and the output matrix are given by (3.34) and (3.35), respectively.

$$\mathsf{Y} = \begin{bmatrix} \mathsf{y}_1^\top \\ \mathsf{y}_2^\top \\ \vdots \\ \mathsf{y}_N^\top \end{bmatrix} = \begin{bmatrix} y_1^1 & y_1^2 & \cdots & y_1^{M-1} & y_1^M \\ y_2^1 & y_2^2 & \cdots & y_2^{M-1} & y_2^M \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_N^1 & y_N^2 & \cdots & y_N^{M-1} & y_N^M \end{bmatrix} \tag{3.33}$$

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \tag{3.34}$$

Since the input signal is considered exactly reproducible in each experiment (i.e. completely deterministic in nature), it is presented by a vector of length N. On account of variability in the system output, a different output time series is produced as a result of each application of the input signal. Each column in this output matrix Y represents the response from each individual experiment. On the other hand, each row represents a value of the response for each experiment at a given instant of time. The output matrix Y can be described with the help of the stochastic theory. Considering any time instant k, the row corresponding to k can be regarded as M realizations of a stochastic process. As this stochastic output y_k is originated from a dynamic stochastic system, the PDF governed by the random output signal y_k is conditioned on the tapped delay lines of the output and input. For the sake of computational tractability, we consider only the envelopes of the response and use them in the proposed modelling. Using the envelopes, the output matrix is reduced to two columns

only, containing the upper and lower bound of the response, respectively. The interval output matrix $\tilde{Y} \in \mathbb{R}^{N\times 2}$ is given as:

$$\tilde{Y} = \begin{bmatrix} \tilde{y}_1^\top \\ \tilde{y}_2^\top \\ \vdots \\ \tilde{y}_N^\top \end{bmatrix} = \begin{bmatrix} \underline{y}_1 & \bar{y}_1 \\ \underline{y}_2 & \bar{y}_2 \\ \vdots & \vdots \\ \underline{y}_N & \bar{y}_N \end{bmatrix} \tag{3.35}$$

where $\tilde{y}_k^\top = [\underline{y}_k, \bar{y}_k]$, $k = 1, \ldots, N$ is the interval output at the k-th instant and $\underline{y}_k$ and $\bar{y}_k$ are its lower and upper bounds, respectively. The description of these envelops are given in the sequel.

3.4.1 The Worst Case Envelopes (Min-Max)

For the worst case envelopes, also referred to as min-max envelopes, the lower and upper bounds of response are taken as the minimum and maximum values of y_k, which are approximated based on the observed minimum $y_k^{\min}$ and maximum $y_k^{\min}$ values as follows:

$$\underline{y}_k = y_k^{\min} = \min_j \left(y_k^j \right), \tag{3.36a}$$

$$\bar{y}_k = y_k^{\max} = \max_j \left(y_k^j \right). \tag{3.36b}$$

Although, the min-max approach is the simplest of its kind to determine the envelopes of the response, it has some serious drawbacks:

- If a model is trained based on the min-max envelopes, it cannot provide the mean or expected response. By calculating the average of the lower and upper bounds of envelopes, we get the mid or centre value of max-min which is not necessarily the expected value of the response.

- The min-max envelopes assume there are no outliers in the data. This makes them least robust statistics in the presence of outliers in data because of their extreme sensitiveness to outliers.

- Using min-max envelopes, the minimum and maximum values of the response are captured for the obtained sample only. Thus, these min-max values are heavily dependent on the given sample especially when the sample size is small.

- The min-max envelopes are conservative.

3.4.2 Percentiles based Envelopes

Percentile based envelopes can be used to overcome some of the drawbacks of the min-max envelopes. In statistics, a percentile is a statistical measure used to indicate a certain level (or score) below which a given percentage of observations falls in a group. For instance, the 5-th percentile is the score below which 5 % of the total observations falls within the sorted list of observations. In this way, those observations which are located at extremities can be avoided, and thus the outliers in data can be handled appropriately. Theoretically, a percentile can be evaluated for any value; however the 25-th percentile (the first quartile), the 50-th percentile (the median), and the 75-th percentile (the third quartile) are widely used. In the literature, there is no standard definition of the percentile. Nevertheless, all those definitions produce similar results as the sample size approaches infinity. In general, the percentile $100 \cdot prc$ of a random variable r is a quantity r_p that satisfies

$$P(\mathsf{r} \leq r_p) = F_{\mathsf{r}}(r_p) = prc, \quad N_o \to \infty, \tag{3.37}$$

where $0 < prc < 1$. $F_{\mathsf{r}}(r_p)$ represents the cumulative distribution function (CDF) of the random variable r evaluated at r_p. The definition is valid in the limit sense, i.e. when the number of observations N_o tends to infinity. The calculation of percentile of data is straightforward. It consists of sorting the data in the ascending order (from smallest to largest) and then assigning the value closest to the percentile of interest. For instance, if we have 1000 data points, the 25th percentile will be the value between 250-th and 251-th sample in the sorted list. Depending upon the definition, the values that fall in between samples should be handled appropriately. For that purpose, approaches like taking the nearest value, calculating the average between two samples, and using linear interpolation are commonly used. In this work, we have used the Matlab function `prctile` [2]. In this method, after sorting the data elements in a data vector, say $\mathbf{x}$, in ascending order, the $100(0.5/N_o)$-th, $100(1.5/N_o)$-th, ..., $100([N_o-0.5]/N_o)$-th percentiles are directly computed, where N_o is the number of elements in $\mathbf{x}$. Given the data points (x_1, y_1) and (x_2, y_2), where $y_1 = f(x_1)$ and $y_2 = f(x_2)$, linear interpolation finds $y = f(x)$ for a given x between x_1 and x_2 as follows:

$$y = f(x) = y_1 + \frac{x - x_1}{x_2 - x_1}(y_2 - y_1), \tag{3.38}$$

where (x_1, y_1) and (x_2, y_2) are the two given data points.

Example 3.4.1. *Consider a data set* $\mathbf{x}$ *consisting of* $N_o = 5$ *elements such as* $\mathbf{x} = \{-2, 2, -1, 0, 1\}$*, the sorted elements* $\{-2, -1, 0, 1, 2\}$ *correspond to the 10-th* $(= 100[0.5/5]$*-th),*

[2] https://de.mathworks.com/help/stats/prctile.html

30-th (= 100[1.5/5]*-th*), *50-th* (= 100[2.5/5]*-th*), *70-th* (= 100[3.5/5]*-th*), *and 90-th* (= 100[4.5/5]*-th*) *percentiles, respectively. See MATLAB help for* [3] *for more detail.*

The approach of determining envelopes using percentiles has the advantage over the min-max approach discussed previously that it excludes outliers in data by considering only a subset of the data vector which lies between two percentiles, say 2.5-th and 97.5-th percentile, and thus avoiding outliers located at extreme values. This approach, however, also has some drawbacks: (1) For large data vectors, sorting elements is computationally expensive, and (2) the best point estimator cannot be determined just using percentiles based envelopes; again the midpoint of the envelopes does not have any significance, it is not equal to the median (50-th percentile) (the same is true for the max-min envelopes).

Example 3.4.2. *This example shown in Fig. 3.7 illustrates the use of min-max and percentile based approaches for estimating the lower and upper bounds or envelopes of 5000 realizations of a random variable* y. *The PDF that* y *follows is given as follows:*

$$\mathsf{y} \sim \begin{cases} \mathcal{N}(-10,1) & \textit{with probability} = 0.01 \\ \mathcal{N}(0,1) & \textit{with probability} = 0.98 \\ \mathcal{N}(10,1) & \textit{with probability} = 0.01 \end{cases} \tag{3.39}$$

where $\mathcal{N}(0,1)$ *is the true or desired normal distribution with zero mean and unity standard deviation. The other two normal distributions, i.e.* $\mathcal{N}(-10,1)$ *and* $\mathcal{N}(10,1)$ *are used to artificially create outliers in* y. *The data is created in such a way that there are 4900 data points from the central lobe of normal distribution, while 100 data points each from the side lobes. Fig. 3.7 clearly indicates that the outliers are very far from the main data coming from the main lobe and thus should not be included for further analysis and processing. Taking the minimum and maximum values of the sample into account includes outliers and provides for unrealistically large bounds. On the other hand, taking the 2.5-th percentile as the lower bound and the 97.5 % percentile as the upper bound of the sample effectively discards the outliers in the data.*

As percentiles approximate a CDF, the equations for output envelopes at each time instant k can be given as:

$$\underline{y}_k = P^{-1}(\mathsf{y}_k < prc^{(low)}\%) = F^{-1}_{\mathsf{y}_k}(prc^{(low)}\%), \tag{3.40a}$$

$$\bar{y}_k = P^{-1}(\mathsf{y}_k < prc^{(upp)}\%) = F^{-1}_{\mathsf{y}_k}(prc^{(upp)}\%), \tag{3.40b}$$

[3] prctile

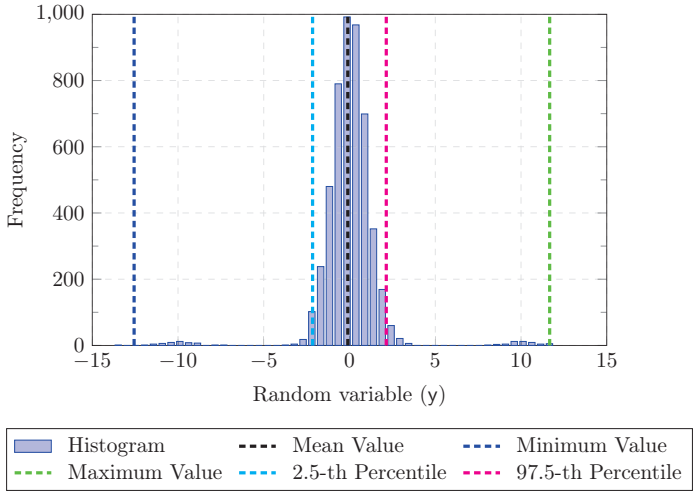


Figure 3.7: Histograms, min-max and percentile based envelopes

where $prc^{(low)}\%$ and $prc^{(upp)}\%$ are, respectively, the lower and upper percentile values in percentage, e.g. $prc^{(low)}\% = 2.5\,\%$ and $prc^{(upp)}\% = 97.5\,\%$ in Example 3.4.2.

3.4.3 Confidence Interval based Envelopes

Definition 3.4.1. *A confidence interval (CI) is a type of an interval estimate of a population parameter given with a probability statement.*

The confidence levels are used to describe the uncertainty associated with the confidence interval estimate. For instance, an interval estimate can be described as "95 % CI". The envelopes which are based on the CIs have some advantages. First, once a coverage factor of the underlying distribution is determined, the calculation of upper and lower bounds becomes straight forward and computationally inexpensive, since it does not require sorting the data as required by the previous approaches. Second, the mean value of the symmetric CIs directly provide the mean or expected value of the random variable. Note that the mean of asymmetric CIs is not equal to the mean or expected value of the random variable distributed asymmetrically. Furthermore, the mean of a skewed distribution provides a non-representative value when a typical or central value is desired. The CI based envelopes

require the specification of the significance level α to calculate the $(1-\alpha)100\,\%$ confidence level. At each time instant, the mean (expected value) and standard deviation (spread) are determined and then the CIs are calculated based on the following formulas:

$$\underline{y}_k = m_{y_k} - k_{\mathrm{cf}} \cdot s_{y_k}, \tag{3.41a}$$

$$\bar{y}_k = m_{y_k} + k_{\mathrm{cf}} \cdot s_{y_k}, \tag{3.41b}$$

where m_{y_k} and s_{y_k} are the mean and standard deviation, respectively, and they are calculated as follows:

$$m_{y_k} = \frac{1}{M} \sum_{j=1}^{M} y_k^j, \tag{3.42}$$

$$s_{y_k} = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} (y_k^j - m_{y_k})^2}. \tag{3.43}$$

k_{cf} is the coverage factor. For a given distribution of known type, this coverage factor is known for getting a desired level of confidence. This term is also known in the literature with the name tolerance critical value and the corresponding confidence interval as the tolerance interval.

Example 3.4.3. *For a normal distribution given by the following PDF*

$$p(\mathsf{r}|\mu_\mathsf{r}, \sigma_r) = \frac{1}{\sigma_\mathsf{r}\sqrt{2\pi}} \exp\left(-\frac{(\mathsf{r}-\mu_\mathsf{r})^2}{2\sigma_\mathsf{r}^2}\right), \tag{3.44}$$

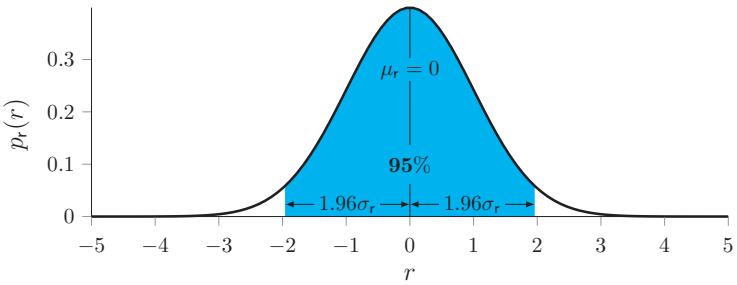
95% of the population lie within the interval determined by the mean of the distribution μ_r and the coverage factor $k_{cf} = 1.96$:

$$\left[\mu_\mathsf{r} - 1.96 \cdot \sigma_\mathsf{r}, \mu_\mathsf{r} + 1.96 \cdot \sigma_\mathsf{r}\right]. \tag{3.45}$$

This is illustrated in Fig. 3.8 for a standard normal distribution with $\mu_\mathsf{r} = 0$ and $\sigma_\mathsf{r} = 1$.

3.4.3.1 Chebyshev's Inequality

In the previous section, it is shown that in order to use the coverage factor, the underlying distribution of data must be known. Knowing the coverage factor, the CI based envelopes can be determined. However, the distribution of data generated from a real world system can be of any arbitrarily shape in the general case. In case of a stochastic dynamic system, such as a system having stochasticity due to friction, the data distribution is quite often



1

Figure 3.8: Illustration of 95% CI of a standard normal distribution with $\mu_r = 0$ and $\sigma_r = 1$

complex. Up to tetra-modal distributions have been observed in the chosen case study of the electro-mechanical throttle as shown in Tab. 3.4. Many existing methods for modelling depend heavily on the unrealistic assumption that the data is normally distributed. Using wrong assumptions for data leads to erroneous and unreliable results. Therefore, we are interested in calculating the coverage factor of any general distribution. The famous Chebyshev's inequality [18] in probability theory provides for the coverage factor for an arbitrary distribution, which is a conservative assessment though.

Definition 3.4.2. *In probability theory, Chebyshev's inequality (also spelled Tchebysheff's inequality) guarantees that in any data sample or probability distribution,"nearly all" values are close to the mean - the precise statement being that no more than* $1/k_c^2$ *of the distribution's values can be more than* k_c *standard deviations away from the mean [99].*

Mathematically, this inequality guarantees that no more than $1/k_{\mathrm{cf}}^2$ of the values of the stochastic output, represented as a random variable y_k, $k = 1, \ldots, N$, of any arbitrary distribution can be farther than $k_{\mathrm{cf}} \geq 1$ standard deviations (σ_{y_k}) apart from the mean (μ_{y_k}), i.e.

$$P(|\mathrm{y}_k - \mu_{\mathrm{y}_k}| \geq k_{\mathrm{cf}}\, \sigma_{\mathrm{y}_k}) \leq \frac{1}{k_{\mathrm{cf}}^2} \tag{3.46}$$

The proof of this inequality is given in Appendix A.1. From this inequality, the coverage factor k_{cf} can be derived easily by equating $1/k_{\mathrm{cf}}^2$ with significance level α:

$$k_{\mathrm{cf}} = \frac{1}{\sqrt{\alpha}} \tag{3.47}$$

Significance level (α)	$(1-\alpha)100\,\%$ CI	$k_{\mathrm{cf}} = \frac{1}{\sqrt{\alpha}}$
0.01	99 %	10.0000
0.05	95 %	4.4721
0.1	90 %	3.1623
0.25	75 %	2.0000
0.5	50 %	1.4142
0.75	25 %	1.1547
0.1	10 %	1.0541

Table 3.3.: The values of the coverage factor (k_{cf}) for the given significance level (α) and the $(1-\alpha)100\%$ confidence interval obtained using Chebyshev's inequality

Example 3.4.4. *In order to determine* $75\,\%$ *confidence interval (*$\alpha = 0.25$*) of* r *such that the probability that values lies outside the interval* $[\mu_{\mathsf{r}} - k_{75\,\%} \cdot \sigma_{\mathsf{r}}, \mu_{\mathsf{r}} + k_{75\,\%} \cdot \sigma_{\mathsf{r}}]$ *must not go beyond 3/4. The desired coverage factor is calculated using* (3.47) *as* $k_{75\,\%} = 1/\sqrt{0.25} = 2$

Tab. 3.3 provides the values of k_{cf} derived from Chebyshev's inequality for some common values of CIs. Theoretically, this inequality can be applied to completely arbitrary distributions; however, for that to be valid, the true value of mean and standard of the distribution are assumed to be known, which contradicts the practical situation, where the distribution and all its parameters are completely unknown in general.

Nevertheless, these true values of the parameters, so-called population parameters in statistics, can be estimated from the unbiased estimates of the sample mean and standard deviation, which are denoted by m_{y_k} and s_{y_k} for the random variable y_k, respectively, provided that the sample size is large enough. Saw et al. [100] first extended Chebyshev's inequality for the case where the population mean and standard deviation are not known, but they are approximated by the sample mean and standard deviation from a finite sample drawn from the same distribution. The resulting extended Chebyshev's inequality is given as:

$$P(|\mathsf{y}_k - m_{y_k}| \geq k_{\mathrm{cf}} \cdot s_{y_k}) \leq \frac{g_{M+1}\left(\frac{M \cdot k_{\mathrm{cf}}^2}{M-1+k_{\mathrm{cf}}^2}\right)}{M+1} \left(\frac{M}{M+1}\right)^{\frac{1}{2}} \tag{3.48}$$

where M is the sample size, m_{y_k} and s_{y_k} are the sample mean and standard deviation, given by (3.42) and (3.43), respectively. The definition of the function $g(r)$ is given as follows: Suppose $r \geq 1$ and $Q = M + 1$, and R be the greatest integer less than Q/r. Denote a^2 as:

$$a^2 = \frac{Q(Q-R)}{1+R(Q-R)}. \tag{3.49}$$

The function $g_Q(r)$ is determined as:

$$g_Q(r) = \begin{cases} R & \text{if } R \text{ is even} \\ R & \text{if } R \text{ is odd and } r < a^2 \\ R-1 & \text{if } R \text{ is odd and } r \geq a^2 \end{cases} \tag{3.50}$$

The inequality (3.48) remains valid when (a) the population moments do not exist and (b) the sample is exchangeably distributed [100]. Typical examples of the exchangeably distributed samples include the cases of i.i.d. random variables and the randomized sampling without replacement from a finite population.

Definition 3.4.3. *An exchangeable sequence of random variables is a finite or infinite sequence* $\mathsf{r}_1, \mathsf{r}_2, \mathsf{r}_3, \ldots$ *of random variables such that for any finite permutation* $\pi(\cdot)$ *of subsets of indices* $1, 2, 3, \ldots$, *the resulting joint probability distribution of the permuted sequence* $\mathsf{r}_{\pi(1)}, \mathsf{r}_{\pi(2)}, \mathsf{r}_{\pi(3)}, \ldots$ *is the same as the point probability distribution of the original sequence [20].*

This inequality is known is Saw-Yang-Mo inequality for finite sample sizes. A table of values for this inequality for sample sizes less than 100 ($M < 100$) has been provided in [50]. The Saw-Yang-Mo inequality was later simplified by Kában [41]. The resulting inequality has the form ([41]):

$$P(|\mathsf{y}_k - m_{y_k}| \geq k_{\text{cf}}\, s_{y_k}) \leq \frac{1}{\sqrt{M(M+1)}} \left(\frac{M-1}{k_{\text{cf}}^2} + 1 \right) \tag{3.51}$$

The approximated lower and upper bounds of y_k, i.e. $\tilde{y}_k^{\top} = [\underline{y}_k, \bar{y}_k]$ based on $(1-\alpha)100\,\%$ CI using the simplified form of Chebyshev's inequality in (3.51) can be calculated using (3.41), where the coverage factor k_{cf} can be derived from (3.51) as follows:

$$k_{\text{cf}} = \sqrt{\frac{M-1}{\alpha\sqrt{M(M+1)} - 1}} \tag{3.52}$$

The derivation of this coverage factor is given in Appendix A.2. As mentioned before, the mean values of y_k are estimated using (3.42).

Although Chebyshev's inequality appears to be extremely effective as it is valid for distributions of any kind, it comes at a price. It generally provides poor or extremely conservative bounds when compared with the cases when the actual distribution is taken into account and its corresponding coverage factor is used.

3.5 Description of the Point Estimator

Mean, median and mode are generally considered as the best point estimates as the measure of central tendency. Since, these concepts have been used in this research to obtain the expected or most likely response from the stochastic dynamic system, they are described briefly in the sequel. A discrete random variable r is characterized by a Probability Mass Function (PMF) $f_{\mathsf{r}}(r)$.

Definition 3.5.1. *A probability mass function (PMF) is a function that gives the probability that a discrete random variable is exactly equal to some value.*

3.5.1 Mean

Definition 3.5.2. *The mean (precisely arithmetic mean) or expected value $\bar{r}_{mean}$ of a discrete random variable r following a certain PMF $f_{\mathsf{r}}(r)$ is the long-run arithmetic average of the random variable r obtained from a finite sample of length M and is evaluated as follows*

$$\bar{r}_{mean} = \sum_{k=1}^{M} r_k f_{\mathsf{r}}(r_k) = \frac{1}{M} \sum_{k=1}^{M} r_k, \tag{3.53}$$

where $f_{\mathsf{r}}(r_k)$ is the value of the PMF at the value r_k.

3.5.2 Median

Definition 3.5.3. *The median (or the 2^{nd} quartile or 50-th percentile) $\bar{r}_{med}$ of a discrete random variable r following a certain PMF $f_{\mathsf{r}}(r)$ is the middle value in the sorted list of the values of a random variable, such that it separates the lower half of the data and satisfies the following inequalities*

$$f_{\mathsf{r}}(r < \bar{r}_{med}) = \frac{1}{2} \quad \textit{and} \quad f_{\mathsf{r}}(r \geq \bar{r}_{med}) = \frac{1}{2} \tag{3.54}$$

3.5.3 Mode

Definition 3.5.4. *The mode (or the most likely value) $\bar{r}_{mod}$ of a discrete random variable* r *following a certain PMF $f_{\mathsf{r}}(r)$ is the value at which the PMF takes its highest value of the discrete probability, i.e.*

$$f_{\mathsf{r}}(r = \bar{r}_{mod})^{4} > f_{\mathsf{r}}(r)|_{\forall r, r \neq \bar{r}_{mod}} \tag{3.55}$$

The values of mean, median and mode are all equal for the case of symmetric distributions, such as the normal distribution. However, for asymmetric distributions, these values do not coincide as shown in the following examples.

Example 3.5.1. *Consider a data set consisting of seven realizations, $\boldsymbol{r} = \{1, 1, 4, 5, 7, 9, 15\}$, of the random variable* r *from a stochastic process. The mean, median and mode are given as $\bar{r}_{mea} = (1 + 1 + 4 + 5 + 7 + 9 + 15)/7 = 6$ (arithmetic average), $\bar{r}_{med} = 5$ (middle value of the sorted list) and $\bar{r}_{mod} = 1$ (most repetitive value), respectively.*

Example 3.5.2. *To illustrate mean, median and mode as the measure of the central tendency of a distribution, we consider a case of an asymmetric bi-modal Gaussian distribution as shown in Fig. 3.9. The random variable* r *is distributed according to the following distribution*

$$\mathsf{r} \sim 0.6 \cdot \mathcal{N}(-0.2, 0.1) + 0.4 \cdot \mathcal{N}(0.2, 0.1)$$

The individual Gaussian components along with the composite are shown in Fig. 3.9. 100 random values were drawn from the distribution and the resulting normalized histogram is also illustrated in the figure. The mean, median and mode of this distribution do not coincide on account of the fact that the distribution is asymmetric. In this example, mode ≤ median ≤ mean.

Therefore, when the task is to capture a typical or central value of the underlying distribution, the appropriate choice among these three characteristic values should be made.

3.5.4 Data for the Point Prediction

Either of the mean, median or mode time series can be used for point prediction. These statistics for measuring central tendency are calculated sample- or row-wise for each value of the time instant k, $k \in \{1, 2, \ldots, N\}$, in the output matrix shown in (3.33). The input-output data pairs for the point prediction are then given as $(\mathbf{u}, \mathbf{y}_{\text{cen}})$, where $\mathbf{u} = (u(1), u(2), \ldots, u(N))$ and $\mathbf{y}_{\text{cen}} = (y_{\text{cen}}(1), y_{\text{cen}}(2), \ldots, y_{\text{cen}}(N))$.

[4] The condition remains valid even if the maximum is not unique

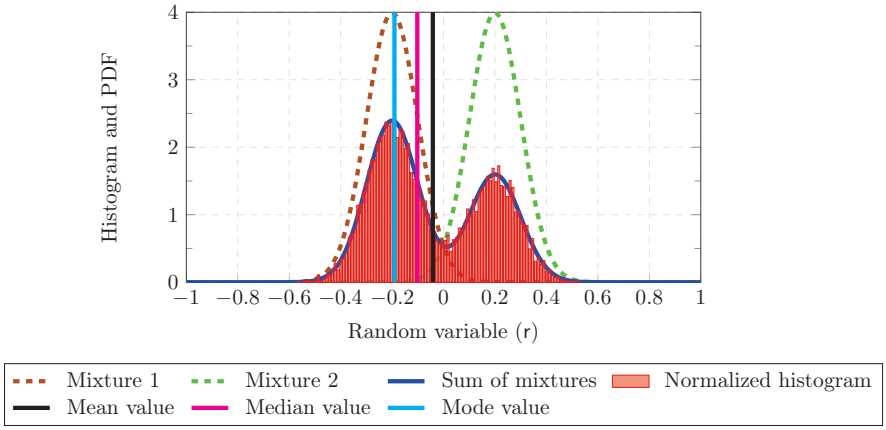


Figure 3.9: Mean, median and mode of an asymmetric bi-modal Gaussian distribution

Example 3.5.3. *If the sample mean is selected as the point estimator, then* $y_{cen}(k) =: y_{exp}(k)$ *is calculated as:*

$$y_{cen}(k) =: y_{exp}(k) = \frac{1}{M}\sum_{j=1}^{M} y_k^j, \quad k = 1, 2, \ldots, N \tag{3.56}$$

3.6 Case Studies

The results of the discussed techniques of uncertainty modelling for the three case studies, namely the simulation academic example, the electro-mechanical throttle, and the servo-pneumatic longitudinal drive will be given next. The general description of these system are given in Chapter 2.

3.6.1 Academic Example

The chosen system is a SISO nonlinear dynamic system with deterministic input and output, as described in Sec. 2.1.1. The deterministic output of this system is given by:

$$y(k+1) = \frac{y(k)y(k-1)(y(k)+2.5)}{1+y^2(k)+y^2(k-1)} + u(k), \tag{3.57}$$

Since we need a stochastic system to be used in the proposed uncertainty modelling approach and later in the system identification, we artificially introduce uncertainty of known type in the output given by (3.57). In order to achieve this task, five random variables following Gaussian distributions with mean and variance given by (3.58) have been added at each time instant k. The resulting output is a mixture of Gaussian distributions (GMM). As a result, the output is made a random variable following a specific and known PDF given as follows:

$$\begin{aligned} \Pr(\mathrm{y}_k) &= 0.5 \cdot \Pr_1 + 0.125 \cdot \Pr_2 + 0.125 \cdot \Pr_3 + 0.125 \cdot \Pr_4 + 0.125 \cdot \Pr_5, \\ \Pr_l &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mathrm{y}_k - \mu_l)^2}{2\sigma^2}\right), \; l = 1, \ldots, 5, \\ \mu_1 &= y(k), \; \mu_2 = 0.9 \cdot y(k), \; \mu_3 = 1.1 \cdot y(k), \mu_4 = 0.8 \cdot y(k), \; \mu_5 = 1.2 \cdot y(k), \; \sigma = 0.1 \end{aligned} \tag{3.58}$$

where $y(k)$ is given by (3.57).

For the sake of illustration, a single time instant for $k = 1000$ is depicted in Fig. 3.10. The red

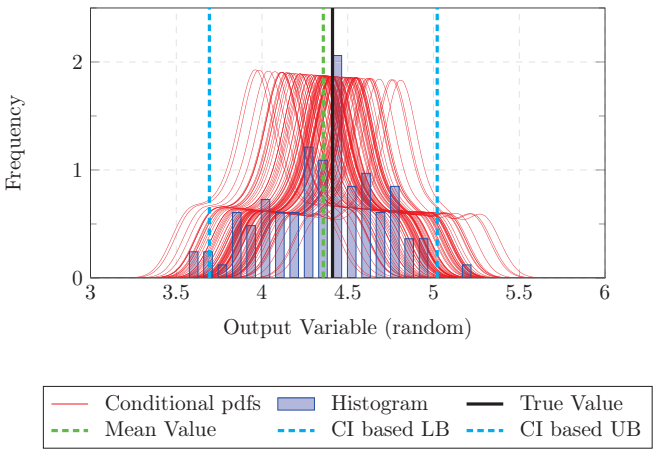


Figure 3.10: y_k at k=1000 [132]

curves show the conditional probability densities $p(\mathrm{y}_k|\mathrm{y}_{k-1}, \ldots, \mathrm{y}_{k-n_y}, u(k-\tau-1), \ldots, u(k-\tau-n_u))$, $k = 1000$. In order to obtain multiple output time series, the experiment is repeated $M = 100$ times. The number of repetitions should not be too low, as it makes the statistical inference unreliable. When it is too high, it increases the experimental burden and the computational overhead. As a rule of thumb, the sample size equals to 25 or 30

considered as the boundary between the small and large sample size [32]. Examples include the central limit theorem, the examination of chi-square distribution, and the reasonably short confidence bounds on the variance estimate for the normal distribution. Therefore, the number of experiments $M = 100$ was chosen as a compromise. There is one curve for the conditional density corresponding to each experiment (shown in red curves). In each of these individual experiments, a random value of output y_k is generated out of this PDF and then subsequently used to determine the conditional density of the next output signal, and so on. At the end, we observe the output matrix, given by (3.33). Recall that, the value of $k = 1000$ actually means the 1000th row of this output matrix. The normalized histogram is also shown in Fig. 3.10 together with the true value of the output signal at that time instant, mean of the observed sample of size 100, and the confidence interval based lower and upper bounds determined using Chebyshev's inequality. The value of the significance level is chosen as $\alpha = 0.25$ for creating envelopes of response based on $100(1-\alpha)100\,\% = 75\,\%$ CI. The corresponding value of the coverage factor is determined using (3.52), using values $\alpha = 0.25$ & $M = 100$, and found out to be $k_{\text{cf}} = 2.0258$.

In order to generate random numbers from any arbitrary distribution in Matlab using the basic rand function (a function that generates a random number from the uniform distribution in the range [0,1]), the approach taken by John S. Denker [23] is used. In this approach, the cumulative density function (CDF) is formed by the cumulative trapezoidal numerical integration of the probability density function. After the CDF is obtained, the inverse CDF function is computed for getting random values of the random variable. For getting the intermediate values between the points, the 1-D interpolation of the inverse CDF is used, see [23] for more details. The whole procedure is illustrated in Fig. 3.11. The lines in Fig. 3.11 show how a uniformly distributed random variable can be mapped into CDF of the desired distribution.

Since the presented modelling approach requires the definition of desired or true point predicted output, the deterministic output given by (3.57) is taken as the reference for point prediction. It is remarked here that in real systems, only a family or band of the output time series is obtained from a stochastic dynamic system and thus the true or reference output time series cannot be exactly defined. However, it is demonstrated that the expected time series is a good estimate of the true time series provided that the uncertainty is distributed symmetrically as given in (3.58) [132]. The output time series of this system in $M = 100$ experiments are illustrated in Fig. 3.12. The figure clearly shows that a band or family of the output time series is obtained as a result of these experiments (shown in grey lines). The red curves shows the envelopes of the response for 75 % CI based on Chebyshev's inequality.

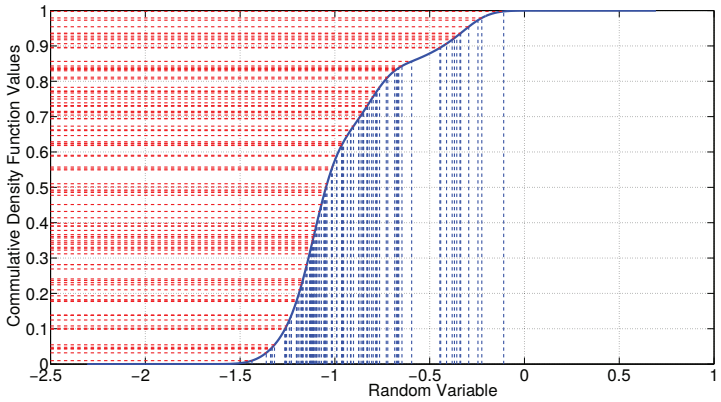


Figure 3.11: Illustration of generation of CDF of a random number following a specific distribution

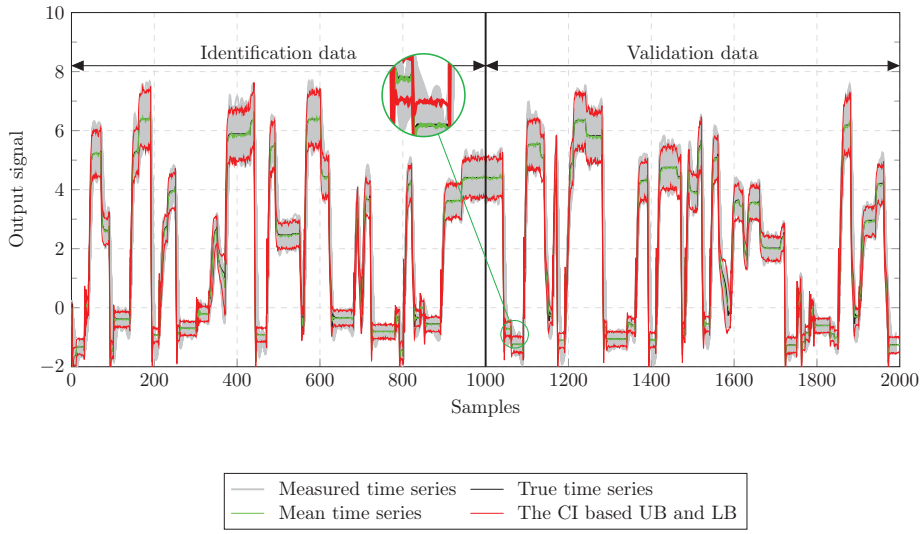


Figure 3.12: The stochastic output signal of the academic example for the identification and validation data [132]

From this figure, it is quite clear that the mean time series is a good estimate of the true time series. Although, only the envelopes based on the confidence interval and mean the

point estimator are illustrated here, similar kind of response is obtained for the worst case and the percentile based envelopes, and median and mode as the point estimators. As described earlier, this scheme is particular useful when the random variable is distributed asymmetrically, since the mean time series then lies exactly in the middle of the confidence interval based envelopes.

3.6.2 Electro-mechanical Throttle Valve

The brief description of the throttle including the motivation behind choosing this as a second test stand, the description of the test stand, and of the design of experiment is given in Sec. 2.2 of Chapter 2. The input signal chosen for carrying out uncertainty analysis is a multisine signal shown in Fig. 3.13. As described in Chapter 2, a phase optimized multisine

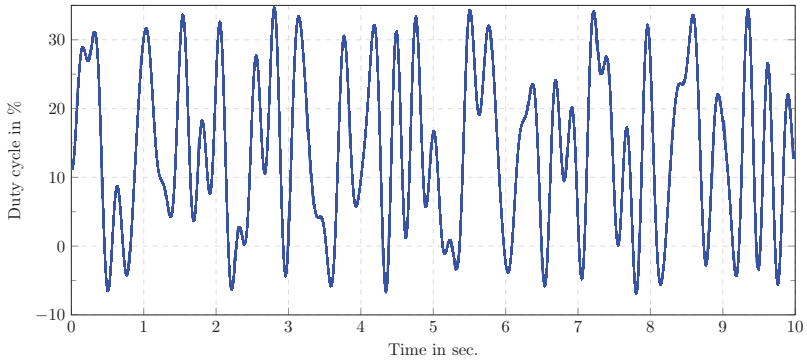


Figure 3.13: The input signal for the electro-mechanical throttle valve for $M = 80$ experiments

signal is selected, and the identification experiment is repeated $M = 80$ times. The number of repetitions should not be too low, as it makes the statistical inference unreliable. When it is too high, it increases the experimental burden and the computational overhead. As a rule of thumb, the sample size equals to 25 or 30 considered as the boundary between the small and large sample size [32]. Examples include the central limit theorem, the examination of chi-square distribution, and the reasonably short confidence bounds on the variance estimate for the normal distribution. Therefore, the number of experiments $M = 80$ was chosen as a compromise. Owing to uncertainties, mainly friction, the output obtained from these experiments shows variability, as depicted in Fig. 3.14. Denoting ϕ as the angular position of the throttle, in all of these experiments the throttle starts from its fully open position, i.e.

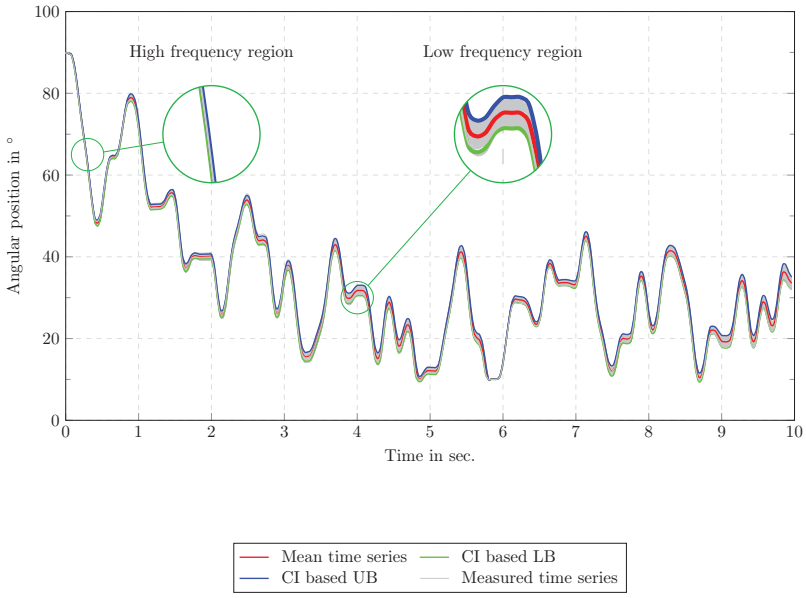


Figure 3.14: The output signal (the angular position of the throttle) for $M = 80$ experiments

$\phi = 90\,°$. A band of output time series is obtained shown in grey and marked as measured time series in Fig. 3.14. Large variability is observed, especially at the slow-changing zone, where the throttle got stuck, where the stick-slip effect is dominant. In the fast-changing zone, for instance the initial duration of 0.4 sec., less variability is observed.

Looking more closely at the distribution of uncertainty at each time instant, it was observed that the uncertainty was distributed non-normally. Furthermore, the form, nature and shape of uncertainty was not uniform at all time instants. For the sake of illustration, the histogram of the output signal at a time instant $k = 100$ is illustrated in Fig. 3.15. To estimate and visually describe the PDF, the normalized histogram, the density estimate by KDE and GMM are illustrated in this figure. As described in Sec. 3.1, increasing the number of Gaussian mixtures or components increases the estimation capability of the GMM. A parsimonious model with reasonable estimation is, however, recommended. Hence, there is always a trade off between the estimation accuracy and the number of parameters in the model. Keeping that trade off in mind, in order to find the adequate number of mixtures components, two criteria AIC and BIC are used. Both the AIC and BIC suggest to use

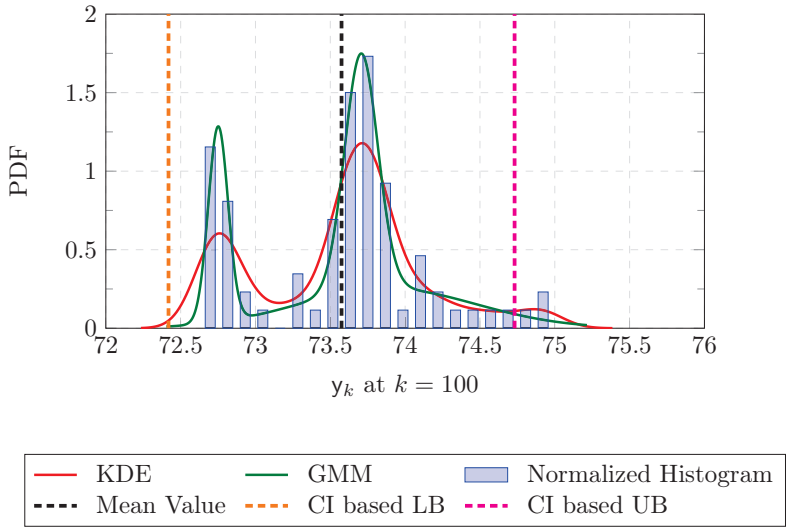


Figure 3.15: Uncertainty analysis of y_k at $k = 100$

No. of mixture	AIC	BIC
1	-1210.57	-1205.36
2	-1250.35	-1237.33
3	-1246.35	-1225.51
4	-1270.18	-1241.53
5	-1264.40	-1227.92
6	-1258.39	-1214.10

Table 3.4.: The values of AIC and BIC for different Gaussian mixtures

four Gaussian mixtures or components in the GMM. The AIC and BIC values are tabulated in Tab. 3.4. The mean and standard deviation of these mixtures along with the mixing coefficient are given in Tab. 3.5.

For this distribution, the extended Chebyshev inequality in (3.48) has been used to determine lower and upper bounds or envelopes of the response based on $(1-\alpha)100\,\%$ CI, where $\alpha = 0.25$ provides for the 75 % CI. For that purpose, the value of the coverage factor k_{cf} is calculated as 2.0324 using (3.52). The mean value along with the envelopes is illustrated by the dotted lines in Fig. 3.15. Since the distribution is asymmetric, the value of y_k at which the PMF takes the maximum value, the so-called most likely value or the mode of the distribution,

Mixture	Mixing coefficient	Mean	Standard deviation
1	0.497	73.703	0.015
2	0.195	72.743	0.003
3	0.217	74.247	0.197
4	0.091	73.056	0.038

Table 3.5.: The parameter values of the GMM

differs from the mean value as indicated in Fig. 3.15. If the most likely value is desired, the distribution has to be considered. However, this would require to train two models. One for the most likely model using crisp data (using the classical TS model), and the other one for envelopes using interval data (TS modelling using the interval data).

Needless to say, for this distribution, actually it does not require to test the distribution for normality, but for the sake of completeness, a normality test is performed on this sample, as described in detail in Sec. 3.2. The data points clearly deviate from the diagonal line, indicating that the underlying distribution is not normal, as shown in Fig. 3.16.

3.6.3 Servo-Pneumatic Longitudinal Drive

For the general description of the servo-pneumatic longitudinal drive see Sec. 2.3. In this section, the uncertainty analysis of the servo-pneumatic longitudinal drive will be done using the analysis tools discussed in this chapter. The analysis is similar to the case study of the throttle valve. As done previously, in order to capture the stochasticity of the output signal at each time instant, we design an input signal and apply it to the system a number of times. For that purpose, a phase optimized multisine is selected as the input signal, see Sec. 2.3 for the design details. The input and the corresponding output signal for $M = 100$ experiments are shown in Fig. 3.17 and Fig. 3.18, respectively. The number of repetitions should not be too low, as it makes the statistical inference unreliable. When it is too high, it increases the experimental burden and the computational overhead. As a rule of thumb, the sample size equals to 25 or 30 considered as the boundary between the small and large sample size [32]. Examples include the central limit theorem, the examination of chi-square distribution, and the reasonably short confidence bounds on the variance estimate for the normal distribution. Therefore, the number of experiments $M = 100$ was chosen as a compromise.

Considerable variability in the output signal is observed due to the stochastic nature of the drive, mainly due to friction. As in the case of the electro-mechanical throttle valve, the experiment is performed in the open-loop and without any friction compensation. Fig. 3.18

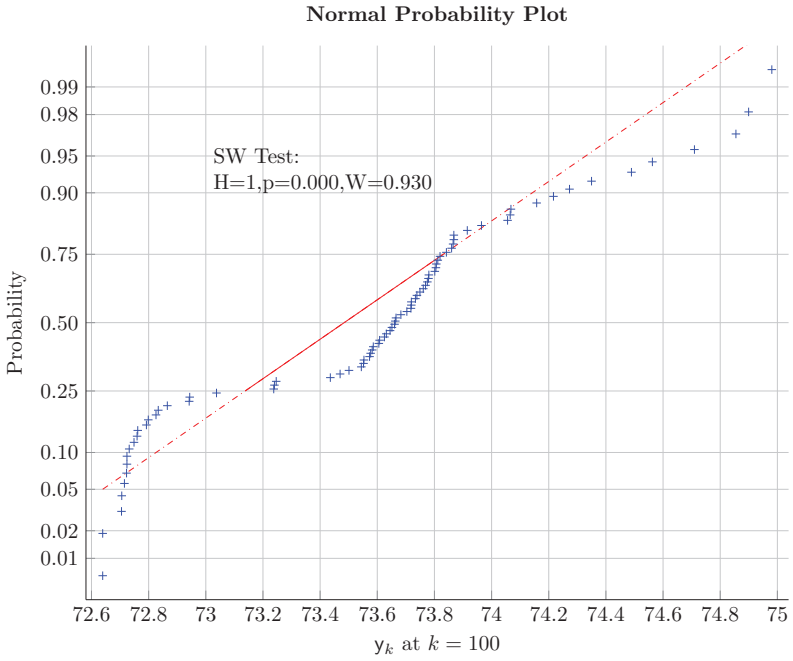


Figure 3.16: Normal probability plot for the chosen sample

shows all the 100 measured output times series along with the mean time series and the envelopes based on the $(1-\alpha)100\,\%$ CI using the extended Chebyshev inequality in (3.48). As in the case of the electro-mechanical throttle, the value of the coverage factor $k_{\text{cf}} = 2.0324$ was calculated using (3.52) for $\alpha = 0.25$ which leads to 75 % CI. As apparent from Fig. 3.18, the effect of uncertainty is prominent and clearly visible at low frequencies, for instance, between 51 and 53 seconds, which is also shown enlarged. The piston at this time instant was not able to move and a different force was required in each experiment to make it move.

In order to perform an in-depth uncertainty analysis at each time instant, the normalized histogram, KDE and GMM are used for density estimation, and it is illustrated for the instant $k = 550$ in Fig. 3.19. Similar to the case of the throttle, up to six mixtures are used to estimate the PDF for GMM. It is remarked here that the distribution shown in Fig. 3.19 is only for one instant. For each time instant, a different distribution is obtained with different

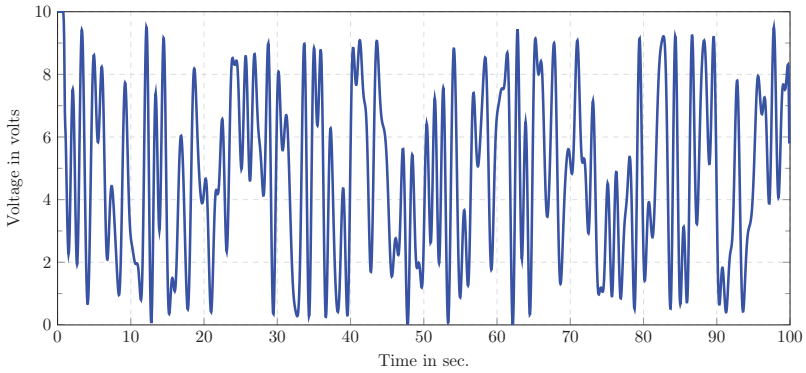


Figure 3.17: The input signal for the servo-pneumatic longitudinal drive for $M = 100$ experiments

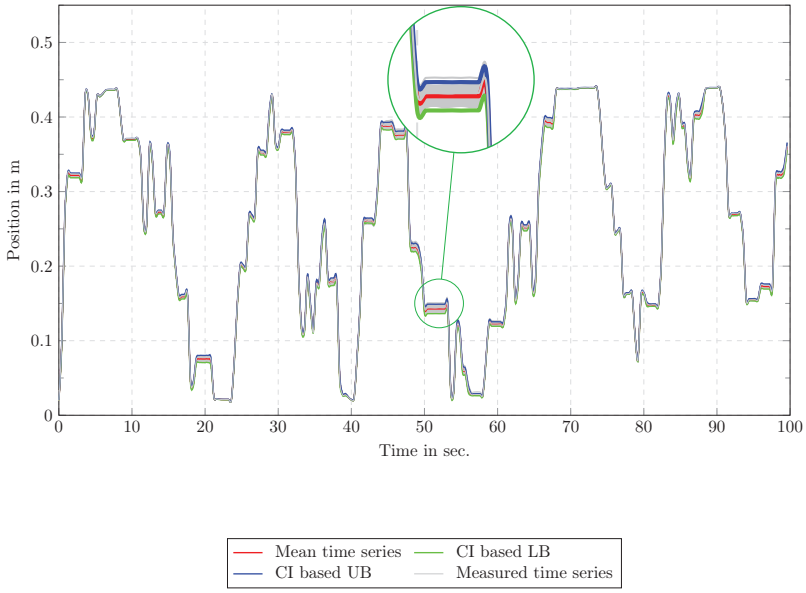


Figure 3.18: The output signal for the servo-pneumatic longitudinal drive for $M = 100$ experiments

values of the standard deviation resulting in different values of the coverage factor. Both the AIC and BIC suggest 3 mixtures to be included in the GMM. The minimum of these values

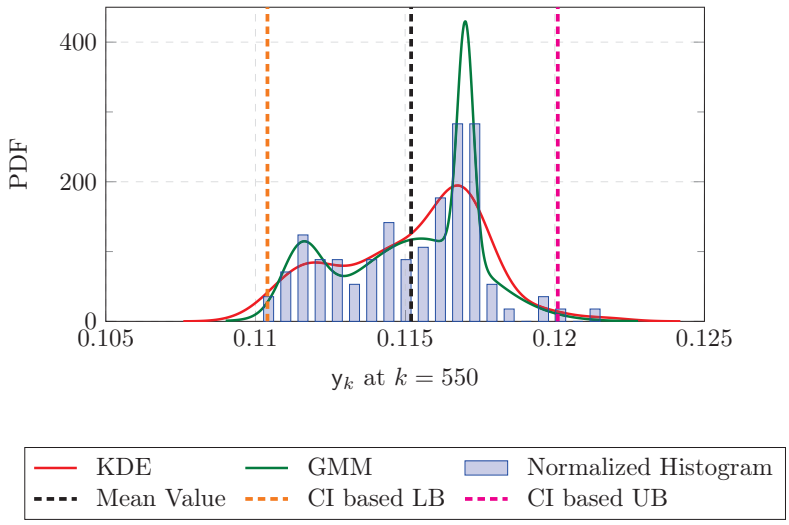


Figure 3.19: The normalized histogram, KDE, GMM, mean value and CI based LB and UB for the output sample of y_k at $k = 550$

Mixture	AIC	BIC
1	-920.21	-915.00
2	-932.59	-919.56
3	-945.97	-925.13
4	-941.90	-913.25
5	-940.67	-904.20
6	-934.64	-890.36

Table 3.6.: The values of AIC and BIC for different Gaussian mixtures

occurred at the third mixture as given in Tab. 3.6. The parameters of GMM are given in Tab. 3.7.

Mixture	Mixing coefficient	Mean	Standard deviation (10^{-7})
1	0.224	0.117	0.70
2	0.160	0.112	4.46
3	0.616	0.116	42.97

Table 3.7.: The parameter values of the GMM

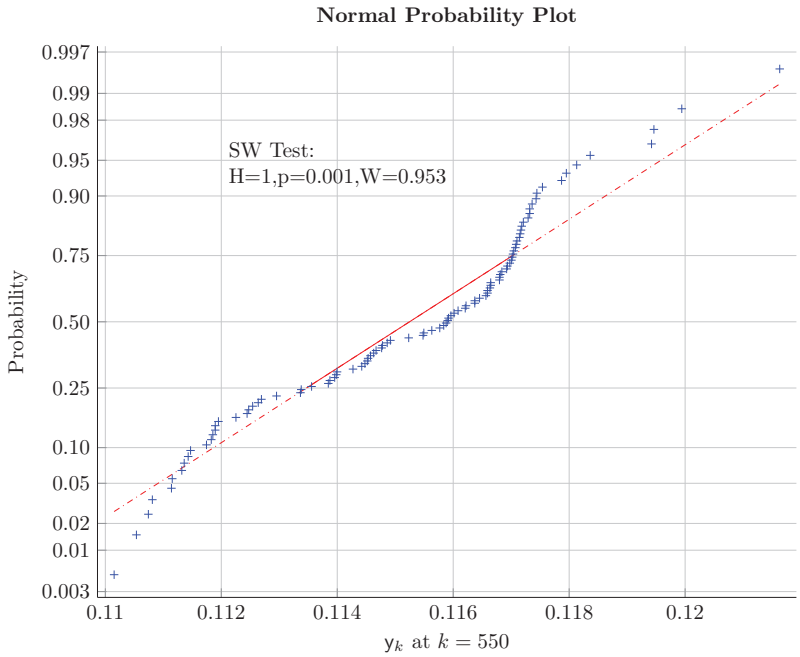


Figure 3.20: Normal Probability Plot

The distribution is certainly non-normal. Its GMM approximation is tri-modal, i.e. it has three Gaussian components. This is also confirmed by the normal probability plot illustrated in Fig. 3.20. The data points do not follow the reference diagonal line for the normal distribution. Looking at the SW test, the normality hypothesis is clearly rejected (H=1) with a fairly low p-value (0.001).

3.7 Summary and Discussion

It is remarked here that the purpose of performing this uncertainty analysis and describing the output values as a histogram is to accurately characterize the uncertainty at each time instant. Since it is not possible to present the analysis for each time instant, only the analysis for a single instance is provided in this chapter. Although, we currently discard this distribution information by taking the envelopes, the modelling can be made more accurate

if the actual distribution, instead of envelopes, is considered for modelling using techniques, such as Symbolic Data Analysis (SDA). This seems to be a tedious task and would require a lot of computational power which is infeasible with the present resources. With the advancement of fast processors and memories, it is expected to process the histogram of the output values at each time instant and thus predicting or forecasting the output signal as a histogram as well.

4

Description of Uncertainties with Fuzzy Logic Systems

This chapter starts with the basic introduction to T2 FLS. The second section provides a succinct overview of the presented modeling approaches in the first section, followed by the definition of some notation that is widely adopted in the forthcoming chapters related to the four types of TS fuzzy modeling. In addition, the model assessment criteria that are used exclusively in all those chapters will be described.

4.1 Introduction to T2 FLS

The conventional FLS use the ordinary two dimensional T1 FS, which have crisp grades. In order to increase the fuzziness of the ordinary FS, T2 FS were proposed which have a fuzzy membership grade. The FLS that use the T2 FS are called T2 FLS. They have proved their higher approximation accuracy and improved uncertainty handling in the Fuzzy System Modelling (FSM). Due to the high computational complexity of the general T2 FLS, the interval T2 FLS (IT2 FLS) were proposed. The interesting and useful characteristic of the T2 FLS is their ability to provide not only the crisp output at the final stage, but also a type-reduced set. The type-reduced set provides a measure of uncertainty in terms of minimum and maximum values of the output, that can be attributed to the presence of the uncertainties in the system. The educational perspective of the T2 FLS as described by Mendel [72] is shown in Fig. 4.1.

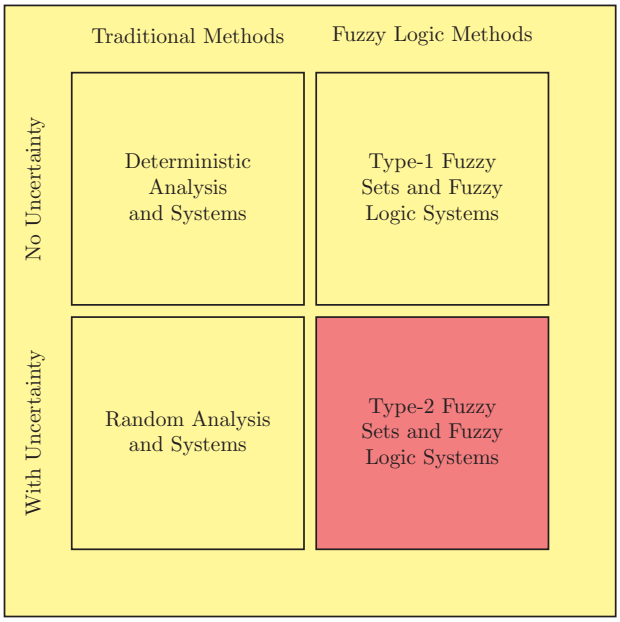


Figure 4.1: Education perspective of T2 FSs and T2 FLSs [72]

Definition 4.1.1. *(T2 FS): A T2 FS, denoted by $\tilde{A}$, is characterized by a T2 MF $\mu_{\tilde{A}}(x,u)$, where $x \in X$ and $u \in J_x \subseteq [0,1]$, i.e.*

$$\tilde{A} = \{((x,u), \mu_{\tilde{A}}(x,u)) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{4.1}$$

in which $0 \leq \mu_{\tilde{A}}(x,u) \leq 1$.

Alternatively, $\tilde{A}$ can also be expressed as:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x,u)/(x,u), \quad J_x \subseteq [0,1] \tag{4.2}$$

Where $\int \int$ denotes union over all admissible x and u in the continuous domain. For discrete universe of discourse $\int$ is replaced by $\sum$. Hence a T2 membership grade can be any subset in $[0,1]$, the primary membership (J_x), and corresponding to each primary membership, there is a secondary membership $(\mu_{\tilde{A}}(x,u))$ that defines the possibilities for the primary membership.

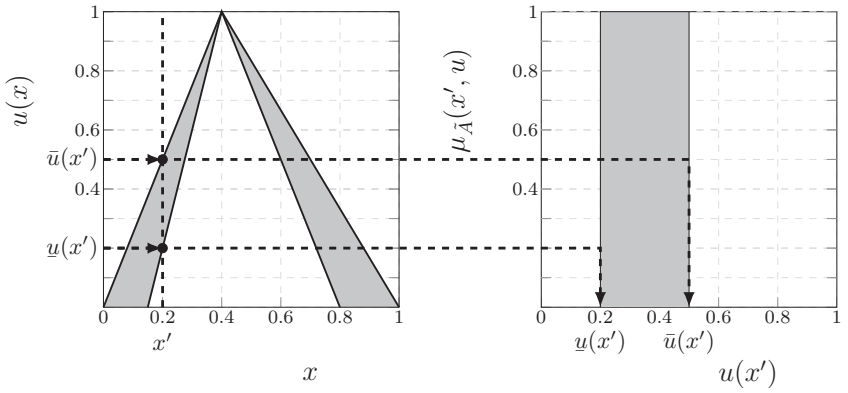


Figure 4.2: An example of IT2 FS (left) and the vertical slice of $x = x'$ (right)

Definition 4.1.2. *(Footprint Of Uncertainty (FOU)): Uncertainty in the primary membership of a T2 FS, $\tilde{A}$, consists of a bounded region that we call the footprint of uncertainty. It is the union of all primary memberships, i.e.,*

$$FOU(\tilde{A}) = \bigcup_{x \in X} J_x \tag{4.3}$$

Definition 4.1.3. *(Upper and lower MF): An upper and lower MF are two T1 MFs that are bounds for the FOU of a T2 FS $\tilde{A}$. The upper MF is associated with the upper bound of the FOU ($\tilde{A}$), and is denoted by $\bar{\mu}_{\tilde{A}}(x)$, $\forall x \in X$. The lower MF is associated with the lower bound of FOU ($\tilde{A}$), and is denoted by $\underline{\mu}_{\tilde{A}}(x)$, $x \in X$, i.e.:*

$$\bar{\mu}_{\tilde{A}}(x) = \overline{FOU}(\tilde{A}) \tag{4.4}$$

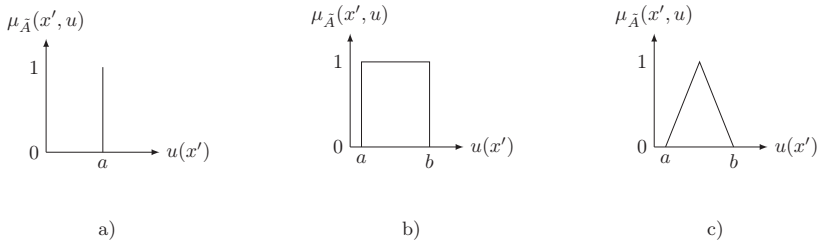
$$\underline{\mu}_{\tilde{A}}(x) = \underline{FOU}(\tilde{A}) \tag{4.5}$$

Definition 4.1.4. *(IT2 FS): An IT2 FS is a special type of T2 FS for which all the secondary grades are equal to unity. It is completely characterized by its FOU. Mathematically,*

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} 1/(x, u), \quad J_x \subseteq [0, 1]. \tag{4.6}$$

In this formula, one can say that for every ordered pair (x, u), the secondary membership grade is unity. An example of IT2 FS is illustrated in Fig. 4.2.

In summary, based on the secondary MF ($\mu_{\tilde{A}}(x,u)$), an FS can be described as T1 FS, IT2 FS and GT2 (General) FS as shown in Fig. 4.3.



a) b) c)

Figure 4.3: Secondary membership of the (a) T1 (single set), (b) interval T2 (interval set), and (c) general T2 (have triangular T1) FSs

The block diagram of a T2 FLS is shown in Fig. 4.4. It consists of a fuzzifier, a data base (FSs), a knowledge base (rules), an inference mechanism, and an output processor. Like the ordinary FLS (T1 FLS), the T2 FLS has the same IF THEN rule structure. Unlike the T1 FLS, the T2 FLS has at least one T2 FS, either in the antecedent or consequent part of the rule and an additional step in the output processing, called type-reduction. Two types of inference mechanism are most popular:

1. Mamdani inference, and
2. Takagi-Sugeno-Kang inference.

Only singleton fuzzification and multi-dimensional TS inference mechanism will be discussed in the later chapters. See [71], for more details of T2 FLS.

The type-reducer converts a T2 FS into its T1 counterpart which can be thought of as representing the uncertainty in the crisp output due to uncertainty in the primary MFs. The details about type-reduction will be presented in the subsequent chapters.

The notation used in this dissertation for describing different types of signals or variables has been organized as shown in Tab. 4.1. Also, an interval variable is denoted by a tilde accent mark (e.g. $\tilde{x}$) in order to distinguish it from a crisp variable (e.g. x). Furthermore, an estimated signal is represented by a hat accent mark. For instance, $\hat{y}$ represents the estimate of y. The lower and upper limits of an interval variable $\tilde{x}$ is denoted by lower and upper bars, respectively, i.e. $\tilde{x} = [\underline{x}, \bar{x}]$.

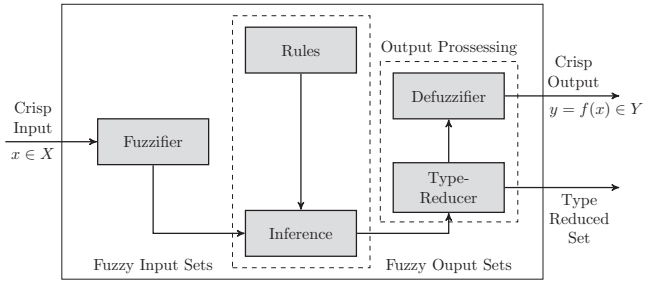


Figure 4.4: Block diagram of T2 FLS

Notation	Example	Description of variable
Roman italic normal lower case	x	Deterministic scalar
Roman italic bold lower case	$\boldsymbol{x}$	Deterministic vector
Roman italic normal upper case	X	Deterministic matrix
Sans-serif normal lower case	x	Stochastic scalar
Sans-serif bold lower case	$\mathbf{x}$	Stochastic vector
Sans-serif normal upper case	X	Stochastic matrix

Table 4.1.: Notation used in this dissertation

4.2 Overview of the Presented Modeling Approaches

In the forthcoming chapters, the T1 FLS with crisp and interval data will be discussed in detail and the results on those models will be demonstrated on the chosen case studies as discussed in Chap. 2. The T1 and IT2 models can be distinguished based on the crisp and interval-valued nature of data and MF values as shown in Fig. 4.5. All of these models can be considered as the extension of the functional fuzzy models using multidimensional reference FSs [52] for interval data and interval memberships. These multidimensional reference FSs

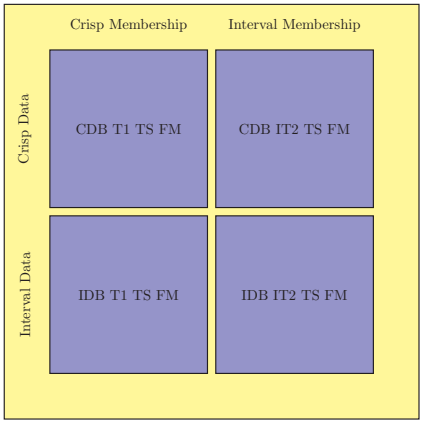


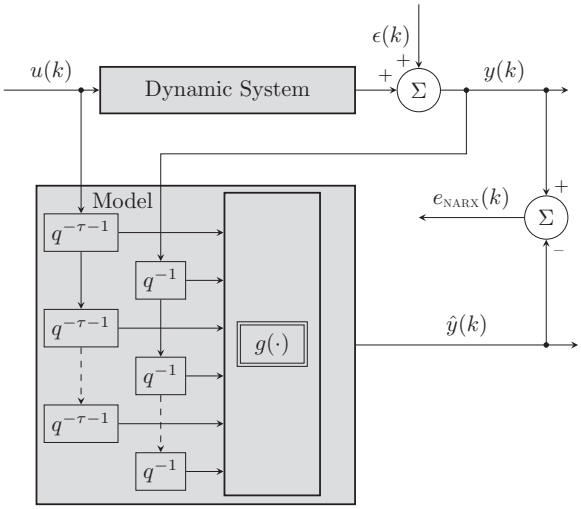
Figure 4.5: The four types of TS fuzzy models based on the crisp and interval nature of data and MF values

were suggested to be used in the modeling of nonlinear systems in [52]. This tremendously reduces the number of rules and thus the problem of curse of dimensionality (especially in the case when there are many inputs) can be avoided by using multidimensional reference FSs.

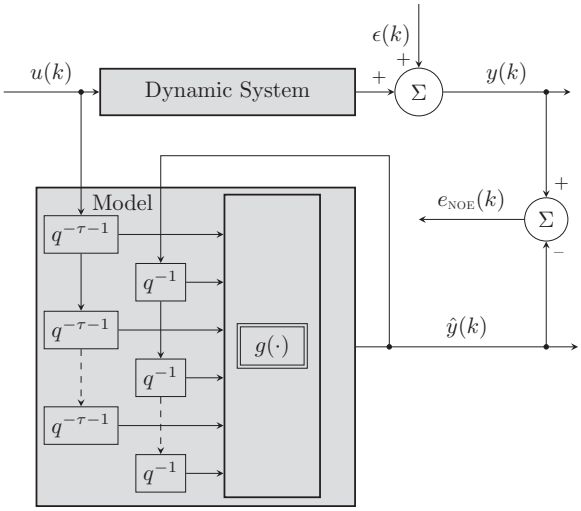
4.3 NARX and NOE Models

Generally, a NARX model is used for prediction, whereas a NOE model is used for simulation [81]. Fig. 4.6 shows the NARX and NOE models. These models consist of tapped delayed blocks (q^{-1}), which are passed to a non-linear function ($g(\cdot)$). In the NARX model, the model predicts one or several steps ahead into the future on the basis of previous process inputs $u(k-\tau-i)$ and process outputs $y(k-i)$ (serial-parallel configuration) and predicted historical output values, and thus the process output needs to be measured during operation. On the other hand, the NOE model simulates future outputs on the basis of previous inputs $u(k-\tau-i)$ only (parallel configuration), and thus process output does not need to be measured during operation. The NARX model can mathematically be described as follows:

$$\hat{y}(k) = g(u(k-\tau-1), \ldots, u(k-\tau-n_u), y(k-1), \ldots, y(k-n_y)) \tag{4.7}$$



a) NARX model (serial-parallel configuration)



b) NOE model (parallel configuration)

Figure 4.6: NARX and NOE models

By the same token, the NOE model can be written as:

$$\hat{y}(k) = g(u(k-\tau-1), \ldots, u(k-\tau-n_u), \hat{y}(k-1), \ldots, \hat{y}(k-n_y)) \tag{4.8}$$

Typical applications for the NARX model include weather forecast and stock market predictions, where the current output of the system is available, or for the cases where an approximate simulation model is required. The NOE model, which is optimized for simulation, is required for applications like accurate simulation or Model Predictive Control (MPC) [38]. Moreover, the NARX model is also used as the initial estimate for the NOE model optimization. It is remarked here that other nonlinear model structures with more complex noise models such as Nonlinear Box-Jenkins models (NBJ) or Nonlinear AutoRegressive Moving Average models with eXogenous inputs (NARMAX) are also possible, but they are uncommon because of their increased computational complexity with increasing number of inputs (curse of dimensionality). Having estimated the NARX model for One Step Ahead Prediction (OSAP), the model is tested for the NOE case. If the modeling performance is not satisfactory, the model parameters need to be adjusted for the NOE model. The problem of estimating the parameters for the NOE model is a nonlinear optimization problem. Knowing the fact that the optimization result highly depends on the initial values[1], the parameters of NARX models are passed to the nonlinear optimization as the best initial guess.

4.4 Model Assessment Criteria

For accessing the performance of a model, usually the approximation error is used as a performance measure. Moreover, the results for the validation/test data set is more significant than the results for the identification/training data set. Among various criteria that have been proposed so far to judge the model quality, usually the measures for the worst case, the average deviation, and/or the relative measure are used. Three criteria, namely Maximum Absolute Error (MaxAE), Root Mean Squared Error (RMSE), and Normalized Mean Squared Error (NMSE) are used in this dissertation. For a complete review see [53]. Given a data set having N observations and assuming $y(k)$ be the output of a system, and $\hat{y}(k)$ be the corresponding output of the model, these criteria are defined as follows:

1. MaxAE:

$$J_{\text{MaxAE}} = \max_{1 \le k \le N} |y(k) - \hat{y}(k)|, \tag{4.9}$$

[1] In the case of gradient-based methods

2. RMSE:

$$J_{\text{RMSE}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y(k)-\hat{y}(k)\right)^2}. \tag{4.10}$$

3. NMSE:

$$J_{\text{NMSE}} = \frac{\sum_{k=1}^{N}(y(k)-\hat{y}(k))^2}{\sum_{k=1}^{N}(y(k)-\bar{y})^2}, \quad \bar{y} = \frac{1}{N}\sum_{k=1}^{N} y(k) \tag{4.11}$$

These criteria are used to asses the performance of the modeling results for both the cases of crisp and interval data. For interval data, the lower and upper bounds of the system and model are first concatenated to form a single time series, and then these criteria are applied.

4.5 Visual Comparison between Alternative Model Descriptions

A visual comparison of the models, with respect to the antecedent and consequent structures and the input and output of the model, is provided in Fig. 4.7. Since only dynamic system identification is addressed in this work, the antecedent ($\mathbf{z}$ or $\tilde{\mathbf{z}}$) and consequent ($\mathbf{x}$ or $\tilde{\mathbf{x}}$) variables consist of lagged inputs and measured (NARX model) or predicted (NOE model) outputs. If a NOE model is considered, an example of these variables is $\mathbf{z}(k) = [u(k-1), \hat{y}(k-1)]^\top$, $\mathbf{x}(k) = [u(k-1), \hat{y}(k-1)]^\top$ for the case of crisp input and crisp output; whereas $\tilde{\mathbf{z}}(k) = [u(k-1), \hat{\tilde{y}}(k-1)]^\top$, $\tilde{\mathbf{x}}(k) = [u(k-1), \hat{\tilde{y}}(k-1)]^\top$ for the case of crisp input and interval output. The input signal shown in Fig. 4.7 is crisp in nature; however, it should not restrict the modeling capability, as the modeling procedure remains the same for an interval input. For the ease of illustration, the MF values are plotted only for the value of 0.5. The value $\nu = 2$ of the fuzziness parameter was selected such that the contour lines of the membership functions (shown in different colors) of each fuzzy set have some distance or gap between them. As $\nu \to 1$, these contour lines seems to coincide with each other. The first of these models is the multi-dimensional reference fuzzy set based fuzzy model presented in [52], and here it is called Crisp-Data Based (CDB) Type-1 (T1) TS Fuzzy Model (FM). This model is called CDB because it uses crisp data (input and output) for model estimation. This is the classical case, in which the model has no uncertainty in the membership functions or the local model parameters. To add the uncertainty handling capability to the classical FM, interval fuzzy membership functions and interval local model parameters are utilized in the CDB Interval Type-2 (IT2) TS FM, resulting in interval bound with bounded uncertainty (so-called typed-reduced set). When there is stochasticity in the dynamic system to be modeled, the output of such a system can be represented in the form of intervals. The last

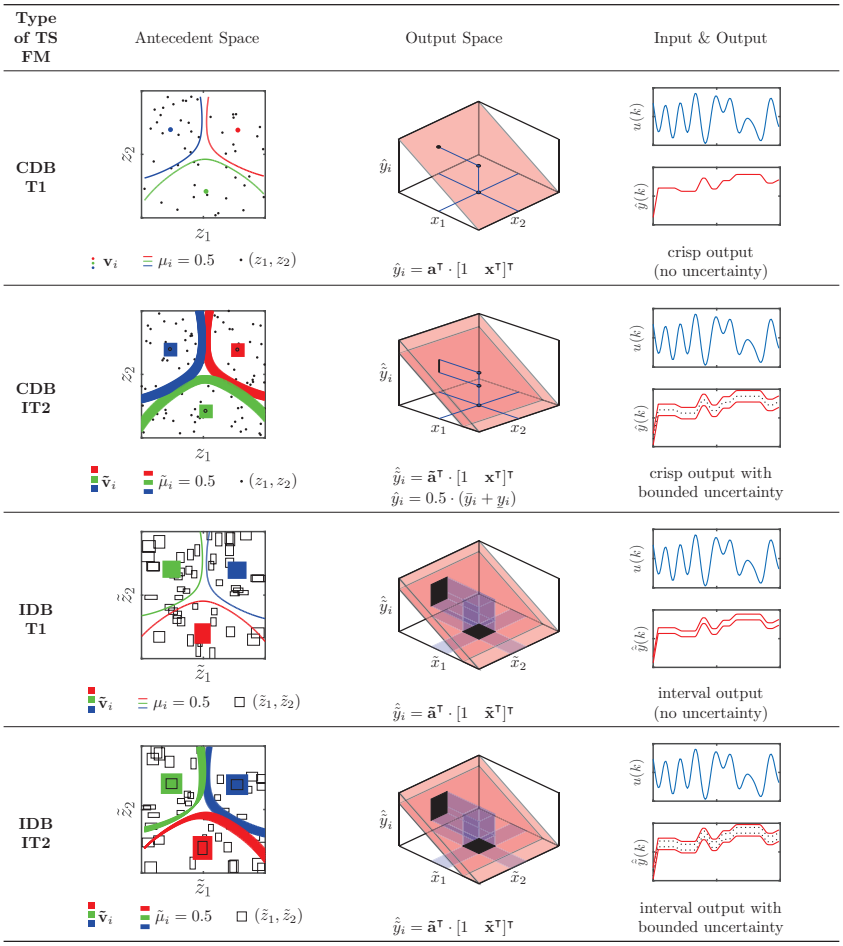


Figure 4.7: Visual comparison between alternative modeling frameworks; CDB T1 FM: Crisp Data-based Type-1 Fuzzy Model, CDB IT2 FM: Crisp Data-based Interval Type-2 Fuzzy Model, IDB T1 FM: Interval Data-based Type-1 Fuzzy Model, IDB IT2 FM: Crisp Data-based Type-1 Fuzzy Model

two models, i.e. the Interval-Data Based (IDB) T1 TS FM, and the IDB IT2 TS FM can be used for modeling such systems using interval data. The difference between these models is summarized in Table 4.2.

Table 4.2.: Comparison of different types of TS FLS

Type of FLS	Interval data $(\hat{\tilde{y}}, \tilde{\mathbf{x}}, \tilde{\mathbf{z}})$	Interval MF $(\tilde{\mu}_i)$	Interval cluster prototypes $(\tilde{\mathbf{v}}_i)$	Interval local model parameters $(\tilde{\mathbf{a}}_i)$
CDB T1 TS	✗	✗	✗	✗
CDB IT2 TS	✗	✓	✓	✓
IDB T1 TS	✓	✗	✓	✓
IDB IT2 TS	✓	✓	✓	✓

4.6 Mathematical Comparisons between Alternative Model Descriptions

Mathematically, these four model types are described for the Multiple-Input Single-Output (MISO) case as follows:

$$\begin{aligned}
\text{CDB T1 TS FM:} \quad & \hat{y}(\mathbf{x},\mathbf{z}) = \sum_{i=1}^{c} \mu_i(\mathbf{v}_i|_{\mathbf{v}_1,\ldots,\mathbf{v}_c}, \mathbf{z}) \left\{ \mathbf{a}_i^\top \begin{bmatrix} 1 & \mathbf{x}^\top \end{bmatrix}^\top \right\}, \\
\text{CDB IT2 TS FM:} \quad & \hat{\tilde{y}}(\mathbf{x},\mathbf{z}) = \sum_{i=1}^{c} \tilde{\mu}_i(\tilde{\mathbf{v}}_i|_{\tilde{\mathbf{v}}_1,\ldots,\tilde{\mathbf{v}}_c}, \mathbf{z}) \left\{ \tilde{\mathbf{a}}_i^\top \begin{bmatrix} 1 & \mathbf{x}^\top \end{bmatrix}^\top \right\}, \\
\text{IDB T1 TS FM:} \quad & \hat{\tilde{y}}(\tilde{\mathbf{x}},\tilde{\mathbf{z}}) = \sum_{i=1}^{c} \mu_i(\tilde{\mathbf{v}}_i|_{\tilde{\mathbf{v}}_1,\ldots,\tilde{\mathbf{v}}_c}, \tilde{\mathbf{z}}) \left\{ \tilde{\mathbf{a}}_i^\top \begin{bmatrix} 1 & \tilde{\mathbf{x}}^\top \end{bmatrix}^\top \right\}, \\
\text{IDB IT2 TS FM:} \quad & \hat{\tilde{y}}(\tilde{\mathbf{x}},\tilde{\mathbf{z}}) = \sum_{i=1}^{c} \tilde{\mu}_i(\tilde{\mathbf{v}}_i|_{\tilde{\mathbf{v}}_1,\ldots,\tilde{\mathbf{v}}_c}, \tilde{\mathbf{z}}) \left\{ \tilde{\mathbf{a}}_i^\top \begin{bmatrix} 1 & \tilde{\mathbf{x}}^\top \end{bmatrix}^\top \right\}.
\end{aligned} \tag{4.12}$$

4.7 Summary and Discussion

Some preliminaries for the rest of the chapters have been presented in this chapter including an overview of the modeling approaches used in this work. Moreover, the notation that is adopted for the fuzzy modeling was described. The model assessment criteria along with their mathematical equations for assessing model quality were described.

The last sections dealt with the visual and mathematical comparison between alternative model descriptions. As can be seen in Fig. 4.7 and is evident from Eq. (4.12), these model descriptions differ from one another depending upon the crisp or interval nature of data or parameters. Of all these models, the IDB IT2 TS FM, is the most general one. It uses interval data for modeling and incorporates bounded uncertainty in interval cluster prototypes and local model parameters.

5

Crisp-Data Based Type-1 TS FLSs

This chapter provides a brief description of the model structure of Crisp-Data Based Type-1 Fuzzy Models (CDB1 T1 FM), along with the procedure to obtain its parameters. Since this model handles crisp input-output data and crisp MF values, it is referred to as the Crisp-Data Based Type-1 Takagi-Sugeno FM. By using the CDB Crisp-Valued (CV) FCM (ordinary FCM) and the method of ordinary least squares, the model is first estimated for the NARX case. Here the ordinary FCM is termed as CDB CV FCM because it clusters the crisp data and provides for crisp cluster prototypes. Subsequently, the NOE model is estimated using a nonlinear optimization algorithm with the NARX parameter values as initial values. Having described the modeling procedure, the modeling results are demonstrated for the chosen case studies. The chapter concludes with a brief discussion.

5.1 Model Structure

To keep things and notations simple, consider a Single-Input-Single-Output (SISO) case. The extension to the Multiple-Input-Single-Output (MISO) case is straightforward, whereas the Multiple-Input-Multiple-Output (MIMO) model can be considered as the extension of several MISO models operating in parallel. The i-th of c rules of the TS FM with multidimensional reference FSs and affine consequents is given by [52]:

$$\text{IF } \mathbf{z} \text{ IS } \mathbf{v}_i \text{ THEN } \hat{y}_i = \mathbf{a}_i^\top \cdot [1, \mathbf{x}^\top]^\top \tag{5.1}$$

with:

R_i:	i-th fuzzy rule,
$\mathbf{z}$:	antecedent or scheduling variable, $\mathbf{z} = [z_1, \ldots, z_{r_a}]^\top \in \mathbb{R}^{r_a \times 1}$,
$\mathbf{v}_i$:	i-th cluster prototype, $\mathbf{v}_i = [v_{1,i}, \ldots, v_{r_a,i}]^\top \in \mathbb{R}^{r_a \times 1}$,
$\hat{y}_i$:	output of the i-th rule, $\hat{y}_i \in \mathbb{R}$,
$\mathbf{a}_i$:	consequent or local model parameters, $\mathbf{a}_i = [a_{0,i}, a_{1,i}, \ldots, a_{r_c,i}]^\top \in \mathbb{R}^{(r_c+1)\times 1}$,
$\mathbf{x}$:	consequent or regressor variable, $\mathbf{x} = [x_1, \ldots, x_{r_c}]^\top \in \mathbb{R}^{r_c \times 1}$.

Since only dynamic systems are considered in this research, $\mathbf{z}$ and $\mathbf{x}$ are chosen as the vectors of lagged inputs and measured outputs in accordance with the NARX nonlinear dynamic model. The antecedent and consequent variables are given by:

$$\mathbf{z}(k) = [u(k-\tau-1), \ldots, u(k-\tau-n_{u_a}), y(k-1), \ldots, y(k-n_{y_a})]^\top \tag{5.2}$$

$$\mathbf{x}(k) = [u(k-\tau-1), \ldots, u(k-\tau-n_{u_c}), y(k-1), \ldots, y(k-n_{y_c})]^\top \tag{5.3}$$

where τ is the input dead time, $r_a = n_{u_a} + n_{y_a}$, and $r_c = n_{u_c} + n_{y_c}$. The values of n_{u_a}, n_{y_a}, n_{u_c}, and n_{y_c} represent the number of lagged input and output terms in the antecedent and consequent variables. For simplicity, the term $\hat{y}_i(\mathbf{x}(k))$ is shortened as $\hat{y}_i(k)$. It is remarked here that, in general, the elements of $\mathbf{z}$ and $\mathbf{x}$ can be any function of lagged inputs and outputs. As an example, the antecedent variable can be selected as: $\mathbf{z}(k) = [u(k-1), y(k-1) - y(k-2)]^T$ with $r_a = 2$ in the electro-mechanical throttle case study. In the multidimensional reference FS, the degree of fulfillment for the i-th rule is determined by the MF of the FCM clustering as follows:

$$\mu_i(\mathbf{z}(k)) = \left[\sum_{j=1}^{c} \left(\frac{\|\mathbf{z}(k) - \mathbf{v}_i\|_2}{\|\mathbf{z}(k) - \mathbf{v}_j\|_2} \right)^{\frac{2}{\nu-1}} \right]^{-1}, \quad \nu \in \mathbb{R}^{>1} \tag{5.4}$$

where $\|.\|_2$ is the Euclidean distance [1], and ν is the fuzziness parameter. The final crisp output of the CDB T1 FLS is calculated as the weighted average of outputs of the c rules, where weights are determined by the MF values as follows:

$$\hat{y}(k) = \sum_{i=1}^{c} \phi_i(\mathbf{z}(k)) \hat{y}_i(\mathbf{x}(k)), \quad \phi_i(\mathbf{z}(k)) = \frac{\mu_i(\mathbf{z}(k))}{\sum_{i=1}^{c} \mu_i(\mathbf{z}(k))} \tag{5.5}$$

where $\{\phi_i(\mathbf{z}(k))\}_{i=1}^{c}$ are called fuzzy basis functions, which actually define the validity regions of the local affine models. The MF of FCM clustering defined by (5.4) are orthogonal, i.e. $\sum_{i=1}^{c} \mu_i(\mathbf{z}(k)) = 1$ and thus $\phi_i(\mathbf{z}(k)) = \mu_i(\mathbf{z}(k))$, $i = 1, \ldots, c$.

[1] It is possible to use other distance norms such as the general Minkowski p-norm (includes Euclidean for $p = 2$) or Mahalanobis distance norm.

The entire identification algorithm comprises of the estimation of the following two sets of parameters:

1. A set of premise or antecedent parameters, i.e. a set of c cluster centers or prototype vectors lumped into a single vector $\mathbf{v} \in \mathbb{R}^{cr_a}$, $\mathbf{v} := [\mathbf{v}_1^\top, \ldots, \mathbf{v}_c^\top]^\top$.

2. A set of consequent or local model parameters consisting of c parameter vectors of local affine models lumped into a single vector $\mathbf{a} \in \mathbb{R}^{c(r_c+1)}$, $\mathbf{a} := [\mathbf{a}_1^\top, \ldots, \mathbf{a}_c^\top]^\top$

The whole procedure to estimate these parameters is summarized in Fig. 5.1.

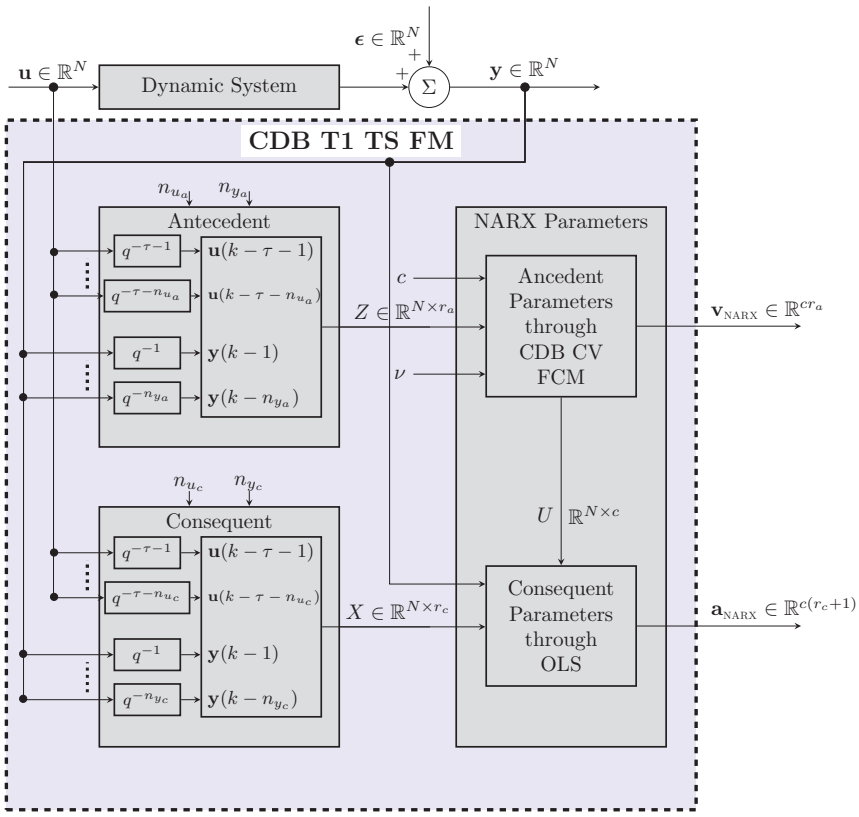


Figure 5.1: The procedure of parameter estimation for CDB T1 TS FM for the NARX case

Having done the DOE, an input signal $\mathbf{u}$ of length N is applied to the dynamic system to be modeled. As a result an output signal $y \in \mathbb{R}^N$ is obtained. These u and y signals are passed to the tapped delay blocks (q^{-1}) of the antecedent (input order = n_{u_a}, output order = n_{y_a}) and consequent (input order = n_{u_c}, output order = n_{y_c}), which lead to the antecedent ($Z \in \mathbb{R}^{N \times r_a}$) and consequent ($X \in \mathbb{R}^{N \times r_c}$) data matrices. The order of the input and output terms for the antecedent and consequent part is chosen a priori by the designer, based on the process knowledge or using some heuristic. These data matrices are then used by the identification algorithm to estimate the antecedent and consequent parameters of the NARX model. By using X, c and ν, the antecedent parameters (cluster prototypes $\mathbf{v} \in \mathbb{R}^{cr_a}$) are estimated through the CDB CV FCM. The consequent parameters ($\mathbf{a}$) are estimated using Ordinary Least Squares (OLS) to minimize the objective function in (5.6):

$$J_{\text{OLS}}^{\text{GBL}} = \sum_{k=1}^{N} \left(y(k) - \sum_{i=1}^{c} \mu_i(\mathbf{z}(k)) \hat{y}_i(k) \right)^2 \stackrel{!}{=} \text{MIN} \tag{5.6}$$

The OLS algorithm provides a consistent, optimal, MLE, and minimum variance unbiased estimate of the parameter vector $\mathbf{a}$, provided that the regressor variables are exogenous (independent of errors), and the errors have finite variance, and are i.i.d. and normally distributed. The parameters obtained by minimizing the objective function (5.6) are optimal in the global sense. When the emphasis is on the interpretation of the local model and the local accuracy, then the parameters should be estimated in the local sense by minimizing (5.7). Here parameters are estimated separately rule by rule.

$$J_{\text{OLS}}^{\text{LCL}} = \sum_{i=1}^{c} \sum_{k=1}^{N} \left(y(k) - \mu_i(\mathbf{z}(k)) \hat{y}_i(k) \right)^2 \stackrel{!}{=} \text{MIN} \tag{5.7}$$

Since the entire procedure of the NARX parameter estimation revolves around the estimation of the antecedent and consequent parameters through the CDB CV FCM clustering and the OLS, respectively, these procedures are given in detail here.

5.2 Partitioning of the Antecedent Space

The Fuzzy C-Means (FCM) [11] is used for the identification of the antecedent/premise. The cluster prototypes ($\mathbf{v}_{\text{NARX}}$) are obtained by minimizing the objective function:

$$\mathbf{v}_{\text{NARX}} := \mathbf{v}^* = \arg\min_{\mathbf{v}} \left(J_{\text{FCM}}(Z, U, \mathbf{v}, \nu) \right), \tag{5.8}$$

with

$$J_{\mathrm{FCM}}(Z, U, \mathbf{v}, \nu) = \sum_{k=1}^{N} \sum_{i=1}^{c} \mu^{\nu}(\mathbf{z}(k)) \|\mathbf{z}(k) - \mathbf{v}_i\|_2^2, \tag{5.9}$$

where Z is the input matrix of the antecedent part,

$$Z := [\mathbf{z}(1), \ldots, \mathbf{z}(N)]^{\top},\ Z \in \mathbb{R}^{N \times r_a},$$

and U is the partition matrix,

$$U := \left[\mu_i(\mathbf{z}(k))\right]_{\substack{i=1,\ldots,c \\ k=1,\ldots,N}},\ U \in \mathbb{R}^{N \times c}.$$

As apparent from the objective function (5.9), the cluster prototypes ($\mathbf{v}_i$) are adjusted in a clustering sense, i.e. with the aim to group the data points. Accordingly, they are not adjusted optimally to capture the input-output behavior of the system.

The implementation of the FCM algorithm begins with the random initialization of either the partition matrix U or the cluster prototypes ($\mathbf{v}_i$), and then it subsequently iterates between the representation and allocation steps. In the representation step, the cluster prototypes are updated as follows:

$$\mathbf{v}_{\mathrm{NARX},i} = \frac{\sum_{k=1}^{N} \mu_i^{\nu}(\mathbf{z}(k)) \mathbf{z}(k)}{\sum_{k=1}^{N} \mu_i^{\nu}(\mathbf{z}(k))} \tag{5.10}$$

In the allocation step, the values of the MF are updated according to:

$$\mu_i(\mathbf{z}(k)) = \left[\sum_{j=1}^{c} \left(\frac{\|\mathbf{z}(k) - \mathbf{v}_i\|_2}{\|\mathbf{z}(k) - \mathbf{v}_j\|_2} \right)^{\frac{2}{\nu - 1}} \right]^{-1}, \quad \nu \in \mathbb{R}^{>1} \tag{5.11}$$

The entire CDB CV FCM clustering algorithm is executed in the following steps:

1: **procedure** CDB CV FCM Algorithm
2: **Initialization:**
3: Choose c (the number of clusters), $2 \leq c < N$
4: Choose ν (fuzziness parameter), $1 < \nu < \infty$
5: Choose T (maximum no. of iterations), $1 \leq T < \infty$
6: Choose ϵ (tolerance limit), $\epsilon > 0$

Randomly initialize $\mathbf{v}_i$ (cluster prototypes), $i = 1, \ldots, c$, $k = 1, \ldots, N$
Set $t = 0$ (iteration counter)
repeat
Representation step:
Fix the membership degrees $\mu_i(\mathbf{z}(k))$'s and compute the prototypes $\mathbf{v}_i$'s, using (5.10)
Allocation step:
Fix the prototypes $\mathbf{v}_i$ and compute the membership degrees $\mu_i(\mathbf{z}(k))$ using (5.11)
$t \leftarrow t + 1$
until $|J_{\text{FCM},t+1} - J_{\text{FCM},t}| < \epsilon$ **or** $t > T$
Output: cluster prototypes ($\mathbf{v}_i$'s) that locally optimize (5.9)
end procedure

5.3 Estimation of the Local Model Parameters

In this work, the consequent parameters ($\mathbf{a}_{\text{NARX}} \in \mathbb{R}^{c(r_c+1)\times 1}$) are estimated globally by using the OLS method (NARX model). Denote $M_i \in \mathbb{R}^{N\times N}$, the diagonal matrix having membership grades $\mu_i(\mathbf{x}(k))$ as its k-th diagonal element with $1 \le i \le c$ and $1 \le k \le N$. Define a matrix

$$X_e := [X, \mathbf{1}] \in \mathbb{R}^{N\times(r_c+1)},$$

where X is the input matrix for the consequent part,

$$X := [\mathbf{x}(1), \ldots, \mathbf{x}(N)]^T \in \mathbb{R}^{N\times r_c},$$

and $\mathbf{1}$ is a unitary column vector in $\mathbb{R}^{N\times 1}$. Moreover, define

$$X' \in \mathbb{R}^{N\times c(r_c+1)}$$

as

$$X' := [M_1 X_e, \ldots, M_c X_e],$$

then $\mathbf{a}_{\text{NARX}}$ is calculated as

$$\mathbf{a}_{\text{NARX}} = [(X')^T X']^{-1} (X')^T \mathbf{y}.$$

The premise and consequent parameters can be lumped into $\boldsymbol{\theta}_{\text{NARX}}$ as follows

$$\boldsymbol{\theta}_{\text{NARX}} \coloneqq [\mathbf{v}_{\text{NARX}}^T, \mathbf{a}_{\text{NARX}}^T]^T \in \mathbb{R}^{c(r_a+r_c+1)\times 1}. \tag{5.12}$$

5.4 Estimation of the NOE Model from the NARX Model

Having estimated the NARX model, its performance is evaluated for the NARX scheme as shown in Fig. 5.2. As the target is to obtain a NOE model, the NARX model is usually first tested for the NOE scheme, shown in Fig. 5.3 When the performance of the NOE

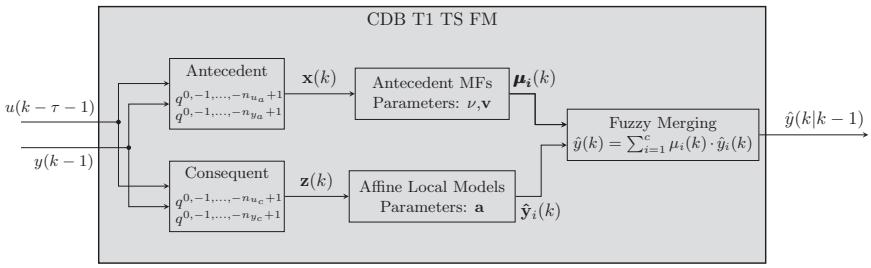


Figure 5.2: The evaluation of the NARX model

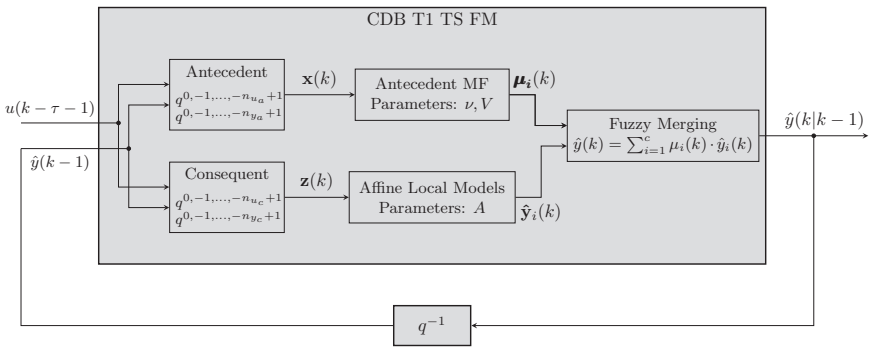


Figure 5.3: The evaluation of the NOE model

scheme is not satisfactory, then the NARX model is optimized for the NOE case. The Matlab function `lsqnonlin` is used for determining optimal cluster prototypes and local

model parameters for parallel mode evaluation. Levenberg-Marquardt [55, 56] is selected as nonlinear optimization algorithm. Denoting the lumped parameter vector for a NOE model as $\boldsymbol{\theta}_{\text{NOE}} \in \mathbb{R}^{c(r_a+r_c+1)\times 1}$, $\boldsymbol{\theta}_{\text{NOE}} := [\mathbf{v}_{\text{NOE}}^\top, \mathbf{a}_{\text{NOE}}^\top]^\top$. It is obtained by minimizing the Mean Squared Error (MSE) of NOE model as follows

$$\boldsymbol{\theta}_{\text{NOE}} := \boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{N}\sum_{k=1}^{N}(y(k) - \hat{y}_{\text{NOE}}(\boldsymbol{\theta}, k))^2. \tag{5.13}$$

The starting value of $\boldsymbol{\theta}$ is chosen to be $\boldsymbol{\theta}_{\text{NARX}}$.

5.5 Case Studies

This section deals with the results of the presented CDB T1 TS FLS on case studies discussed in Chap. 2. For the general description of these systems along with the mathematical formulation and DOE, see Chap. 2.

5.5.1 Academic Example

The uniformly distributed random input signal chosen for this artificial academic example is shown in Chap. 2. The data is split into two halves; the first half is for identification and the other half is for validation. The clustering is performed in the input space and the value of fuzziness parameter is chosen as $\nu = 1.3$. Moreover, the number of local models is selected as $c = 6$ as a compromise between model complexity (parsimony) and approximation accuracy. It is demonstrated in Fig. 5.4, where the number of local models are plotted against the NMSE of the NOE model for the test dataset. The following antecedent and consequent variables are used: $\mathbf{z}(k) = \mathbf{x}(k) = [u(k-1), y(k-1), y(k-2)]^\top$. The Sequential Forward Selection (SFS) [27, 89] method was used for selecting these variables, in which each input is selected sequentially to optimize the total squared error. The output signal together with the output of the NARX model is illustrated in Fig. 5.5. As expected the output of the NARX model looks very accurate. The NARX model is then evaluated for the NOE case. The results for the NOE model are further improved using nonlinear optimization as described in Sec. 5.4 and demonstrated in Fig. 5.6. The modeling results for the academic case study are given in Tab. 5.1.

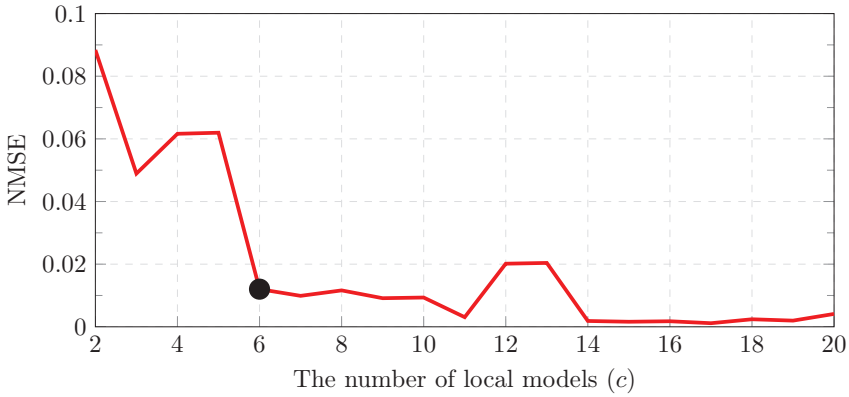


Figure 5.4: The number of local models (c) versus the NMSE of the NOE model for the test dataset

Type of CDB T1 TS FLS	Identification Data			Validation Data		
	MaxAE	NMSE ($\times 10^{-3}$)	RMSE	MaxAE	NMSE ($\times 10^{-3}$)	RMSE
NARX	1.119	3.395	0.154	1.175	3.110	0.148
NOE_WO	2.198	8.271	0.240	2.202	12.048	0.290
NOE_OPT	1.015	1.959	0.117	0.664	2.065	0.120

Table 5.1.: Values of MaxAE, NMSE and RMSE for the NARX, NOE_WO (Without Optimization) and NOE_OPT (Optimized) models for the academic case study

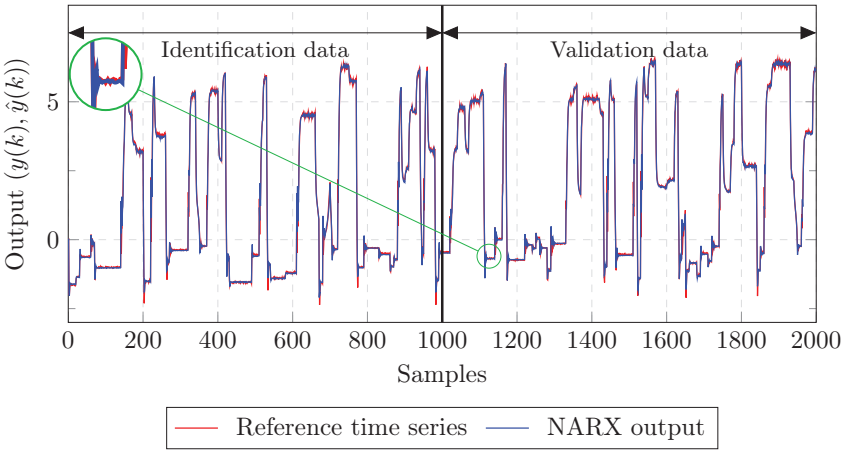


Figure 5.5: The true output time series and the output of the NARX model for the academic example

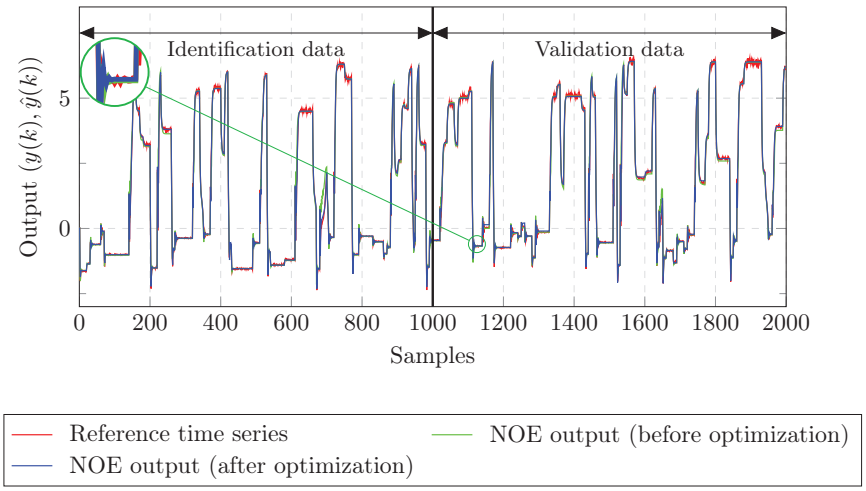


Figure 5.6: The true output time series and the output of the NOE model before and after optimization for the academic example

5.5.2 Electro-mechanical Throttle Valve

The phase optimized multisine signal chosen for the electro-mechanical throttle is shown in Chap. 2. The clustering is performed in the input space, as the results of clustering in the product space does not change the result. The value of fuzziness parameter is chosen as ν = 1.3. Moreover, the number of local models is selected as c = 8 as a compromise between model complexity and approximation accuracy. It is demonstrated in Fig. 5.7, where the number of local models are plotted against the NMSE of the NOE model for the test dataset. The

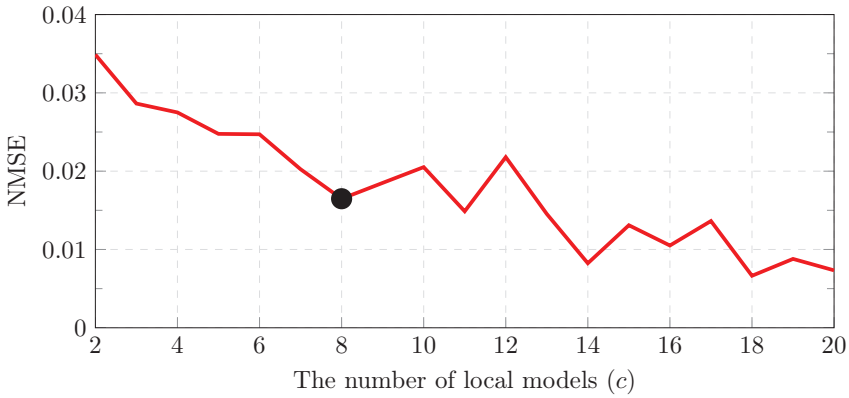


Figure 5.7: The number of local models (c) versus the NMSE of the NOE model for the test dataset

following antecedent and consequent variables are used: $\mathbf{z}(k) = [u(k-1), y(k-1)-y(k-2)]^\top$, $\mathbf{x}(k) = [u(k-1), y(k-1), y(k-2)]^\top$. As previously, they are selected based on the Sequential Forward Selection (SFS) [27, 89] method. The sampling time is chosen to be T_s = 10 ms. The first 900 (90 %) data points are used for identification, whereas the remaining 100 (10 %)data points are used for validation. This split was based on the experience about the system. There should be enough data points for identification, in order to obtain the desired modelling performance. Reducing the amount of data in the identification data set, reduces the modelling performance. The model is first trained for the NARX case and the results obtained are illustrated in Fig. 5.8. Having estimated the NARX model, the NOE model was evaluated. The modeling results of the NOE model before and after optimization are demonstrated in Fig. 5.9. The results of the modeling for the electro-mechanical throttle case study are given in Tab. 5.2. As apparent from Tab. 5.2, the modelling performance of

Type of CDB T1 TS FLS	Identification Data			Validation Data		
	MaxAE in °	NMSE ($\times 10^{-3}$)	RMSE in °	MaxAE in °	NMSE ($\times 10^{-3}$)	RMSE in °
NARX	0.914	0.001	0.055	0.110	0.034	0.033
NOE_WO	4.4688	15.17	2.144	1.250	20.253	0.817
NOE_OPT	0.486	0.058	0.133	0.398	1.232	0.210

Table 5.2.: Values of MaxAE, NMSE and RMSE for the NARX, NOE_WO (Without Optimization) and NOE_OPT (Optimized) models for the electro-mechanical throttle case study

the NOE model improved significantly after optimization, which shows the importance of using optimization for this case study.

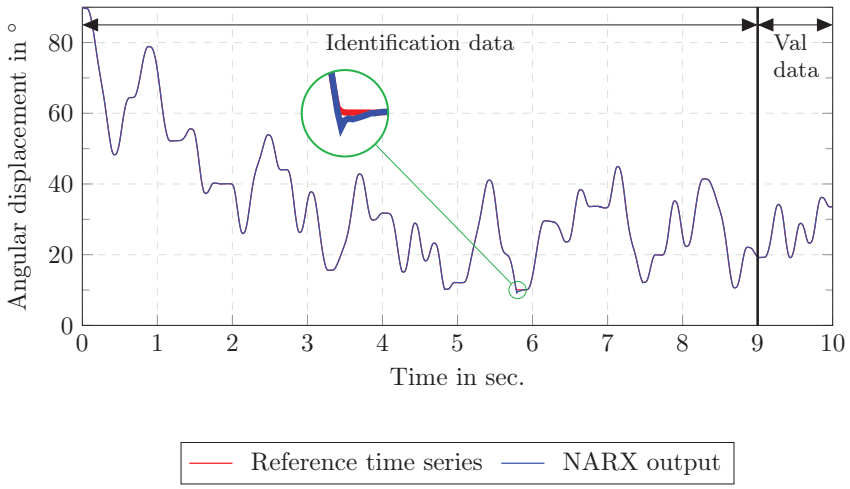


Figure 5.8: The results of the NARX model for the electro-mechanical throttle

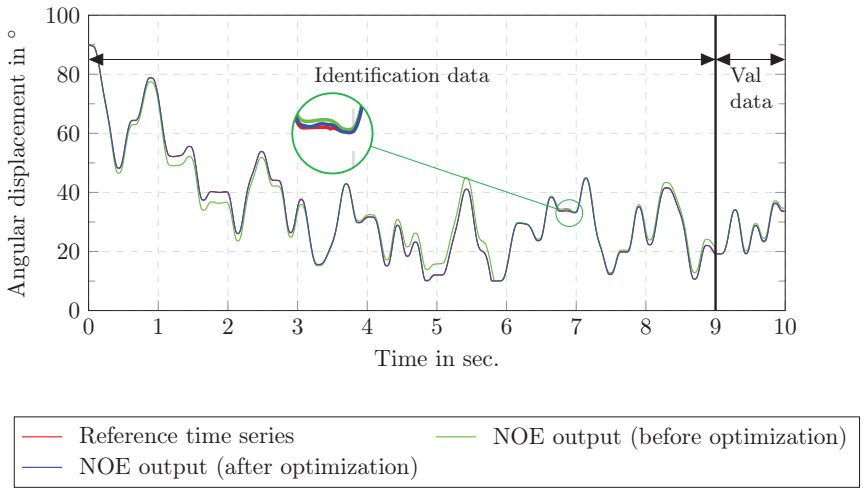


Figure 5.9: Measured output time series and output of the NOE model before and after optimization for the electro-mechanical throttle

5.5.3 Servo-Pneumatic Longitudinal Drive

The phase optimized multisine signal chosen for the servo-pneumatic longitudinal drive is shown in Chap. 2. For this case study, the value of the fuzziness parameter $\nu = 1.1$, and the number of local models $c = 10$ is selected. The value of the number of local models is selected as a compromise between model complexity and approximation accuracy as demonstrated in Fig. 5.10. Like the previous two case studies, the clustering is performed in the input space. The following antecedent and consequent variables are used: $\mathbf{z}(k) = [u(k-1), y(k-1), y(k-2)]^\top$, $\mathbf{x}(k) = [u(k-1), u(k-2), u(k-3), y(k-1), y(k-2), y(k-3)]^\top$. It is a usual practice to have the same antecedent and consequent variables. However, since the Sequential Forward Selection (SFS) [27, 89] method is applied separately on both the antecedent and consequent variables, different antecedent and consequent variables were obtained. The sampling time is chosen as $T_s = 0.1$ s. The first 700 data points are used for identification, whereas the remaining 100 data points are used validation. This split was based on the experience about the system and the system constraints. Reducing the number of data points to below 700 resulted in the deteriorated system performance, and thus this split was selected. The data points for the validation data set can be increased by increasing the length of the experiment. However, it is not done in this work. The model is first trained for the NARX case and the

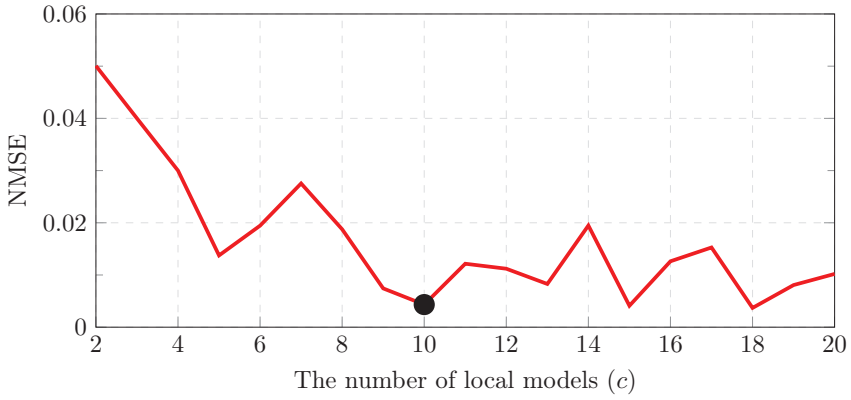


Figure 5.10: The number of local models (c) versus the NMSE of the NOE model for the test dataset

results obtained are illustrated in Fig. 5.11. Finally, the modeling results of the NOE model before and after optimization are demonstrated in Fig. 5.12.

Type of CDB T1 TS FLS	Identification Data			Validation Data		
	MaxAE in m	NMSE ($\times 10^{-3}$)	RMSE in m	MaxAE in m	NMSE ($\times 10^{-3}$)	RMSE in m
NARX	0.026	1.368	0.005	0.021	1.167	0.004
NOE_WO	0.079	71.048	0.036	0.026	6.392	0.010
NOE_OPT	0.021	1.601	0.005	0.030	7.1555	0.011

Table 5.3.: Values of MaxAE, NMSE and RMSE for the NARX, NOE_WO (Without Optimization) and NOE_OPT (Optimized) models in the servo-pneumatic longitudinal drive case study

The results of the modeling for the servo-pneumatic longitudinal drive case study are given in Tab. 5.3. Note that, after the optimization of the NOE model, the NMSE value slightly decreases, whereas the MaxAE and RMSE values more or less remain the same. The performance on the identification has improved significantly, however.

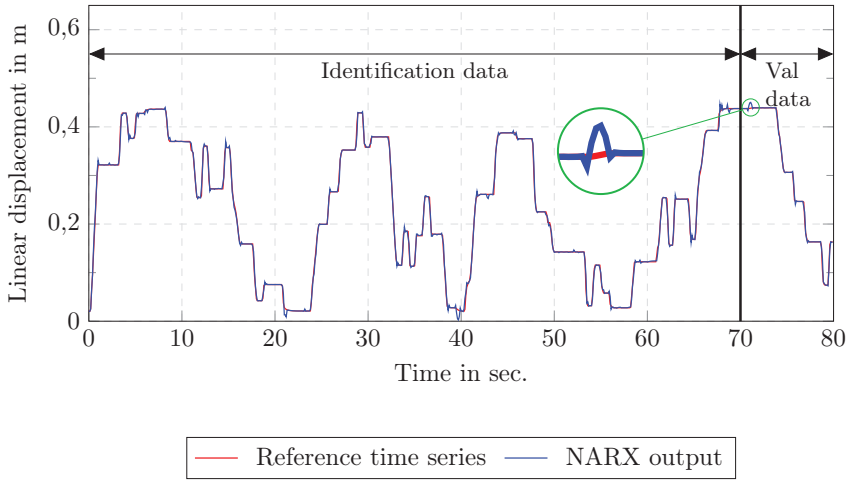


Figure 5.11: The results of the NARX model for the servo-pneumatic longitudinal drive

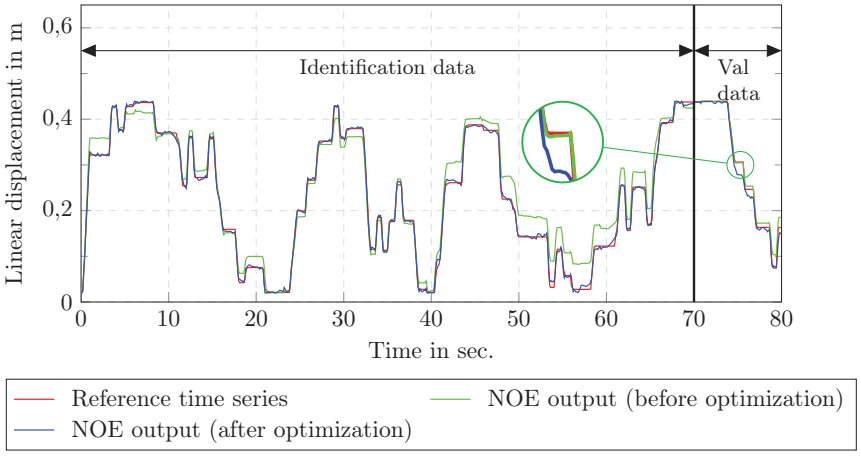


Figure 5.12: The measured output time series and the output of the NOE model before and after optimization for the servo-pneumatic longitudinal drive

5.6 Summary and Discussion

The modeling of dynamic systems using CVD T1 TS FLS has been described in this chapter. Owing to the universal approximation capability of the Takagi-Sugeno fuzzy models, they are widely used to model non-linear dynamic systems. By selecting the appropriate model structure and parameters, it is possible to capture the system dynamics using the input-output data. The results for the three case studies demonstrated that the presented modeling methodology is capable to capture the crisp-valued output time series, which is actually the expected output response. The modelling procedure starts with the estimation of the NARX model. The fuzzy c-means clustering was used for determining the antecedent structure, whereas the ordinary least squares method was used to estimate the parameters of the local model. The model was succeedingly tested for the NOE case. In order to improve the performance of the NOE model, the parameters of the NOE model were optimized by using the non-linear optimization algorithm.

In the sequel, modeling results will be presented using interval data to capture the envelopes of response resulting from the stochastic nature of the system (more specifically, the variability in the system output). The modeling results can be further improved by choosing an optimal input signal, increasing the signal length, tuning the parameters of the nonlinear optimization algorithm, and the selection of optimal model structure etc., which is, however, not the focus of this research.

6

Interval-Data Based Type-1 TS FLSs

A concise description of the model structure of the IDB T1 FM together with the procedure to obtain the model parameters is provided in this chapter. This model uses interval-data for identification and has the type-1 fuzzy sets in the antecedent part. By using the so-called Interval-Data Based Crisp-Valued FCM clustering (IDB CV FCM) and the method of linear interval regression, a NARX model is first estimated followed by the estimation of the NOE model using a nonlinear optimization algorithm with the NARX parameter as the initial estimate. The modeling results of IDB TS FM are demonstrated in the chosen case studies. In the end, some concluding remarks are given in the discussion section.

6.1 Model Structure

The i-th rule of the IDB T1 TS FM is characterized by a multivariate fuzzy set having interval-valued prototypes and a set of interval consequent parameters and is given by:

$$R_i: \quad \text{IF } \tilde{\mathbf{z}} \text{ IS } \tilde{\mathbf{v}}_i \text{ THEN } \hat{\tilde{y}}_i(\tilde{\mathbf{x}}) = \tilde{\mathbf{a}}_i^{\top} \begin{bmatrix} 1 & \tilde{\mathbf{x}}^{\top} \end{bmatrix}^{\top} \tag{6.1}$$

where:

R_i: i-th fuzzy rule,
$\tilde{\mathbf{z}}$: interval type antecedent or scheduling variable, $\tilde{\mathbf{z}} = [\underline{\mathbf{z}}, \bar{\mathbf{z}}] \in \mathbb{R}^{r_a \times 2}$,
$\tilde{\mathbf{v}}_i$: interval type i-th cluster prototype, $\tilde{\mathbf{v}}_i = [\underline{\mathbf{v}}_i, \bar{\mathbf{v}}_i] \in \mathbb{R}^{r_a \times 2}$,
$\tilde{\hat{y}}_i$: interval type output of the i-th rule, $\tilde{\hat{y}}_i = [\underline{\hat{y}}_i, \bar{\hat{y}}_i] \in \mathbb{R}^{1 \times 2}$,
$\tilde{\mathbf{a}}_i$: interval type consequent or local model parameters, $\tilde{\mathbf{a}}_i = [\underline{\mathbf{a}}_i, \bar{\mathbf{a}}_i] \in \mathbb{R}^{(r_c+1) \times 2}$,
$\tilde{\mathbf{x}}$: interval type consequent or regressor variable, $\tilde{\mathbf{x}} = [\underline{\mathbf{x}}, \bar{\mathbf{x}}] \in \mathbb{R}^{r_c \times 2}$.

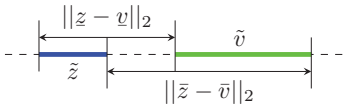
The degree of fulfillment of the premise of the i-th rule is determined by the membership function of the IDB CV FCM clustering. Mathematically:

$$\mu_i(\tilde{\mathbf{z}}(k)) = \left[\sum_{h=1}^{c} \left(\frac{d^2_{\text{INT}}(\tilde{\mathbf{z}}(k), \tilde{\mathbf{v}}_i)}{d^2_{\text{INT}}(\tilde{\mathbf{z}}(k), \tilde{\mathbf{v}}_h)} \right)^{\frac{1}{\nu-1}} \right]^{-1}, \tag{6.2}$$

where $d^2_{\text{INT}}(\cdot,\cdot)$ is the squared Euclidean distance between the interval type vectors, $\nu \in \mathbb{R}^{>1}$ is the fuzziness parameter, and $\mu_i(\tilde{\mathbf{z}}) \in [0,1]$ is the crisp membership value of the interval scheduling variable ($\tilde{\mathbf{z}}$) to the i-th cluster. Owing to the orthogonality of the membership functions $\sum_{i=1}^{c} \mu_i(\tilde{\mathbf{z}}) = 1$ holds. Given the two interval vectors $\tilde{\mathbf{z}} = [\underline{\mathbf{z}}, \bar{\mathbf{z}}] \in \mathbb{R}^{r_a \times 2}$, and $\tilde{\mathbf{v}}_i = [\underline{\mathbf{v}}_i, \bar{\mathbf{v}}_i] \in \mathbb{R}^{r_a \times 2}$, the squared Euclidean distance [1] between them is defined by:

$$d^2_{\text{INT}}(\tilde{\mathbf{z}}, \tilde{\mathbf{v}}_i) = \|\underline{\mathbf{z}} - \underline{\mathbf{v}}_i\|_2^2 + \|\bar{\mathbf{z}} - \bar{\mathbf{v}}_i\|_2^2 \tag{6.3}$$

This is graphically illustrated in Fig. 6.1. Note that, although both the scheduling variable



$$d^2_{\text{INT}}(\tilde{z}, \tilde{v}) = ||\underline{z} - \underline{v}||_2^2 + ||\bar{z} - \bar{v}||_2^2$$

Figure 6.1: The squared Euclidean distance d^2_{INT} between two intervals $\tilde{z}$ and $\tilde{v}$

and the cluster prototypes are interval variables, the distance between them is a crisp number as given in Eq. (6.3), which leads to T1 fuzzy sets. More about this clustering is given in the next section. The output of the IDB T1 TS FM is calculated as the weighted composition

[1] Note that other distance norms defined for interval numbers can also be used such as the Hausdorff distance norm between intervals.

of the interval output of each local model, where weights are given by their corresponding crisp membership values as follows:

$$\hat{\tilde{y}}(\tilde{\mathbf{x}}, \tilde{\mathbf{z}}) = \sum_{i=1}^{c} \mu_i(\tilde{\mathbf{z}}) \cdot \hat{\tilde{y}}_i(\tilde{\mathbf{x}}), \tag{6.4}$$

The entire identification algorithm comprises of the estimation of the following two sets of parameters:

1. A set of interval premise or antecedent parameters, i.e. a set of c cluster centers or prototype vectors lumped into a single vector $\tilde{\mathbf{v}} = [\underline{\mathbf{v}}, \bar{\mathbf{v}}] \in \mathbb{R}^{cr_a \times 2}$, $\underline{\mathbf{v}} := [\underline{\mathbf{v}}_1^\top, \ldots, \underline{\mathbf{v}}_c^\top]^\top \in \mathbb{R}^{cr_a}$, $\bar{\mathbf{v}} := [\bar{\mathbf{v}}_1^\top, \ldots, \bar{\mathbf{v}}_c^\top]^\top \in \mathbb{R}^{cr_a}$.

2. A set of interval consequent or local model parameters consisting of c parameter vectors of local affine models lumped into a single vector $\tilde{\mathbf{a}} = [\underline{\mathbf{a}}, \bar{\mathbf{a}}] \in \mathbb{R}^{c(r_c+1)\times 2}$, $\underline{\mathbf{a}} := [\mathbf{a}_1^\top, \ldots, \mathbf{a}_c^\top]^\top$

Note that for programming convenience in Matlab, the sets of antecedent and consequent vectors are lumped into vectors, rather than matrices. The whole procedure to estimate these parameters is summarized in Fig. 6.2.

The input signal $\mathbf{u} \in \mathbb{R}^N$ is applied to the system M times and leads to M output time series $\mathbf{y}$ each having the length of N samples. The additive measurement noise term is represented by $\mathcal{E}$ which is generally i.i.d. zero mean and finite variance; however, the proposed approach does not require any assumption about $\mathcal{E}$. The pre-processing converts the output time series into interval form, i.e. it reduces the M columns of the output matrix $Y \in \mathbb{R}^{N\times M}$ to two columns for the lower ($\underline{\mathbf{y}} \in \mathbf{R}^N$) and upper ($\bar{\mathbf{y}} \in \mathbb{R}^N$) output time series or envelops. The details of this step are provided in Chap. 3. In order to estimate the NARX model parameters of the IDB T1 TS FM, the interval output time series $\tilde{\mathbf{y}} = [\underline{\mathbf{y}}, \bar{\mathbf{y}}]$ is first converted to antecedent ($\underline{Z}, \bar{Z} \in \mathbb{R}^{N\times r_a}$) and consequent $\underline{X}, \bar{X} \in \mathbb{R}^{N\times r_c}$ matrices, which take their corresponding orders of input and output lagged terms as inputs. These matrices are defined as follows:

$$\underline{Z} = [\mathbf{u}(k-\tau-1), \ldots, \mathbf{u}(k-\tau-n_{u_a}), \underline{\mathbf{y}}(k-1), \ldots, \underline{\mathbf{y}}(k-n_{y_a})] \tag{6.5}$$

$$\bar{Z} = [\mathbf{u}(k-\tau-1), \ldots, \mathbf{u}(k-\tau-n_{u_a}), \bar{\mathbf{y}}(k-1), \ldots, \bar{\mathbf{y}}(k-n_{y_a})] \tag{6.6}$$

$$\underline{X} = [\mathbf{u}(k-\tau-1), \ldots, \mathbf{u}(k-\tau-n_{u_c}), \underline{\mathbf{y}}(k-1), \ldots, \underline{\mathbf{y}}(k-n_{y_c})] \tag{6.7}$$

$$\bar{X} = [\mathbf{u}(k-\tau-1), \ldots, \mathbf{u}(k-\tau-n_{u_c}), \bar{\mathbf{y}}(k-1), \ldots, \bar{\mathbf{y}}(k-n_{y_c})] \tag{6.8}$$

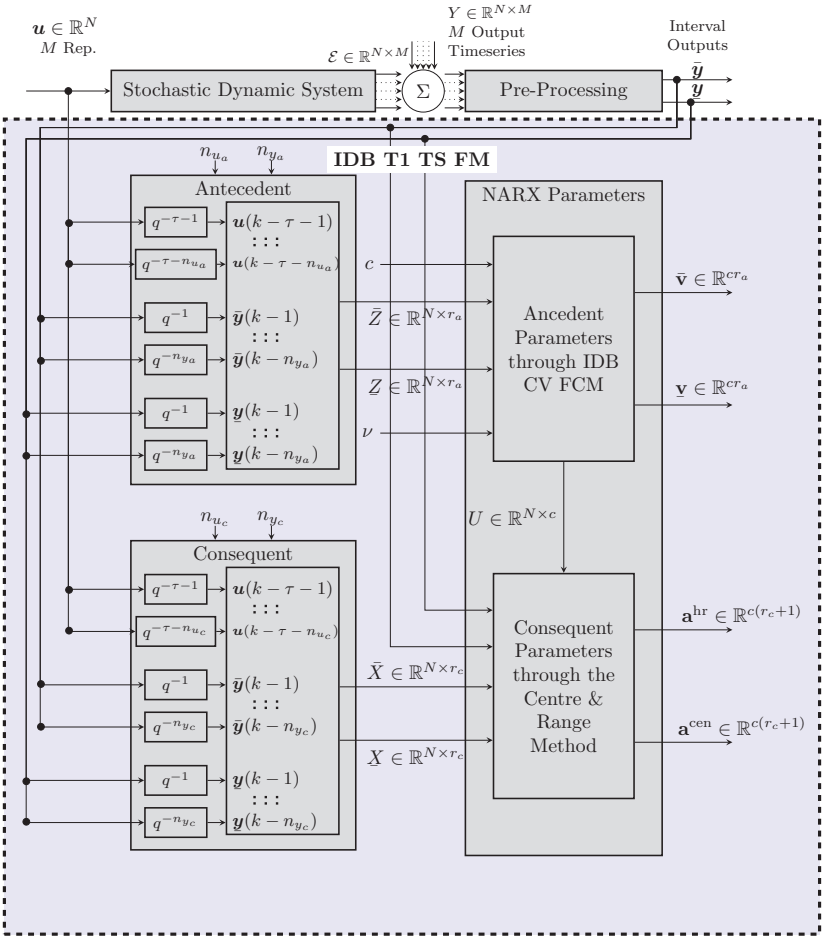


Figure 6.2: The block diagram illustrates the procedure of parameter estimation for IDB T1 TS FM for the NARX case

From these matrices, the interval vectors can be formed by a simple concatenation:

$$\tilde{\mathbf{z}}(k) = [\underline{\mathbf{z}}(k), \bar{\mathbf{z}}(k)] = \begin{bmatrix} \tilde{z}_1(k) \\ \dots \\ \tilde{z}_{r_a}(k) \end{bmatrix} = \begin{bmatrix} u(k-\tau-1) & u(k-\tau-1) \\ \dots & \dots \\ u(k-\tau-n_{u_a}) & u(k-\tau-n_{u_a}) \\ \underline{y}(k-1) & \bar{y}(k-1) \\ \dots & \dots \\ \underline{y}(k-n_{y_a}) & \bar{y}(k-n_{y_a}) \end{bmatrix} \in \mathbb{R}^{r_a \times 2}, k = 1, \dots, N \tag{6.9}$$

$$\tilde{\mathbf{x}}(k) = [\underline{\mathbf{x}}(k), \bar{\mathbf{x}}(k)] = \begin{bmatrix} \tilde{x}_1(k) \\ \dots \\ \tilde{x}_{r_c}(k) \end{bmatrix} = \begin{bmatrix} u(k-\tau-1) & u(k-\tau-1) \\ \dots & \dots \\ u(k-\tau-n_{u_a}) & u(k-\tau-n_{u_c}) \\ \underline{y}(k-1) & \bar{y}(k-1) \\ \dots & \dots \\ \underline{y}(k-n_{y_c}) & \bar{y}(k-n_{y_c}) \end{bmatrix} \in \mathbb{R}^{r_c \times 2}, k = 1, \dots, N \tag{6.10}$$

These interval vectors are then used to estimate the antecedent and consequent parameters in the realm of symbolic interval-valued data as described in the following sections.

6.2 Partitioning of the Antecedent Space

The scheduling variable $\tilde{\mathbf{z}}(k)$, which consists of both the point-valued data (input $\mathbf{u}$) and the interval-valued data (output $\tilde{\mathbf{y}}$), is clustered using the Interval-Data Based Crisp-Valued Fuzzy C-Mean (IDB CV FCM) clustering algorithm. Considering the fact that point-valued data can be seen as a special case of interval-valued data, $\tilde{\mathbf{z}}(k)$ is dealt with the theory of symbolic interval-valued data. The IDB CV FCM clustering furnishes the fuzzy partitioning of the space of $\tilde{\mathbf{z}}(k)$ and provides interval prototypes ($\tilde{\mathbf{v}}$). A brief description of this method is given next, see [1] and the references therein for details.

Let the interval-valued data to be clustered be given by $\tilde{\mathbf{z}}(k) = [\tilde{z}_1(k), \dots, \tilde{z}_{r_a}(k)]^\top$, $k = 1, \dots, N$; where $\tilde{z}_j(k) = [a_k^j, b_k^j] \in \mathfrak{I}_{a,b} = \{[a,b] : a, b \in \mathbb{R}, a \leq b\}$, $j = 1, 2, \dots, r_a$. Let the prototype $\tilde{\mathbf{v}}_i$ of each cluster $\tilde{P}_i$ be represented as a vector of intervals, i.e. $\tilde{\mathbf{v}}_i = [\tilde{v}_1^i, \dots, \tilde{v}_{r_a}^i]^\top$, $i = 1, \dots, c$, where $\tilde{v}_j^i = [\alpha_i^j, \beta_i^j] \in \mathfrak{I}_{\alpha,\beta} = \{[\alpha, \beta] : \alpha, \beta \in \mathbb{R}, \alpha \leq \beta\}$, $j = 1, 2, \dots, r_a$. Let $\nu \in \mathbb{R}^{>1}$ be the fuzziness parameter. The IDB CV FCM algorithm minimizes (locally) the adequacy criterion based on squared Euclidean distances between vectors of intervals as follows:

$$J_{\text{IDB-CV-FCM}}(\tilde{Z}, U, \tilde{v}, \nu) = \sum_{i=1}^{c} \sum_{k=1}^{N} \mu_i^{\nu}(\tilde{\mathbf{z}}(k)) \sum_{j=1}^{r_a} \left[(a_k^j - \alpha_i^j)^2 + (b_k^j - \beta_i^j)^2 \right]. \tag{6.11}$$

As the standard CDB CV FCM, the IDB CV FCM algorithm starts with the random initialization of the cluster prototypes, and then subsequently iterates between the representation and allocation steps. In the representation step, the clustering prototypes are updated as follows:

$$\alpha_i^j = \frac{\sum_{k=1}^{N} \mu_i^{\nu}(\tilde{\mathbf{z}}(k)) a_k^j}{\sum_{k=1}^{N} \mu_i^{\nu}(\tilde{\mathbf{z}}(k))} \quad \text{and} \quad \beta_i^j = \frac{\sum_{k=1}^{N} \mu_i^{\nu}(\tilde{\mathbf{z}}(k)) b_k^j}{\sum_{k=1}^{N} \mu_i^{\nu}(\tilde{\mathbf{z}}(k))}. \tag{6.12}$$

In the allocation step, the values of the memberships are updated according to:

$$\mu_i(\tilde{\mathbf{z}}(k)) = \left[\sum_{h=1}^{c} \left(\frac{\sum_{j=1}^{r_a} \left[(a_k^j - \alpha_i^j)^2 + (b_k^j - \beta_i^j)^2 \right]}{\sum_{j=1}^{r_a} \left[(a_k^j - \alpha_h^j)^2 + (b_k^j - \beta_h^j)^2 \right]} \right)^{\frac{1}{\nu-1}} \right]^{-1}. \tag{6.13}$$

The IDB CV FCM clustering algorithm for symbolic interval data is executed in the following steps [1]:

1: **procedure** IFCM Algorithm
2: **Initialization:**
3: Choose c (Number of clusters), $2 \leq c < N$
4: Choose ν (fuzziness parameter), $1 < \nu < \infty$
5: Choose T (maximum no. of iterations), $1 \leq T < \infty$
6: Choose ϵ (tolerance limit), $\epsilon > 0$
7: Randomly initialize $\tilde{v}_i^j$ (interval cluster prototypes), $i = 1, \ldots, c, j = 1, \ldots, r_a$
8: Set $t = 0$ (iteration counter)
9: **repeat**
10: **Representation step:**
11: Fix the membership degrees $\mu_{i,k}$ and compute the prototypes $\tilde{v}_i^j$ using (6.12)
12: **Allocation step:**
13: Fix the prototypes $\tilde{v}_i^j$ and compute the membership degrees $\mu_{i,k}$ using (6.13)
14: $t \leftarrow t + 1$
15: **until** $|W_{t+1} - W_t| < \epsilon$ **or** $t > T$

16: **Output:** cluster prototypes ($\tilde{\mathbf{v}}_i$) that locally optimize (6.11)
17: **end procedure**

6.3 Estimation of the Local Model Parameters

The center and range method [64] is used for the estimation of the local model parameters because of its simplicity and the fact that the parameters of the local models are estimated in a least-squares optimal sense. The method applies the weighted linear regression on midpoints (centers) and ranges of the interval valued consequent variable $\tilde{\mathbf{x}}(k)$, which can later be used for determining the output $\hat{\tilde{y}}(k)$.

Let the consequent variable $\tilde{\mathbf{x}}(k)$ be written as $\tilde{\mathbf{x}}(k) = [\tilde{x}_1(k), \ldots, \tilde{x}_{r_c}(k)]^\top$ with $\tilde{x}_j(k) = [c_k^j, d_k^j] \in \mathfrak{I}_{c,d} = \{[c,d] : c,d \in \mathbb{R}, c \leq d\}$, $j = 1, \ldots, r_c$ and $k = 1, \ldots, N$. Further assume that $\mathbf{x}^{\text{cen}}(k) = [x_1^{\text{cen}}(k), \ldots, x_{r_c}^{\text{cen}}(k)]^\top$, where $x_j^{\text{cen}}(k) = 0.5(c_k^j + d_k^j)$ and $\mathbf{x}^{\text{hr}}(k) = [x_1^{\text{hr}}(k), \ldots, x_{r_c}^{\text{hr}}(k)]^\top$, where $x_j^{\text{hr}}(k) = 0.5(d_k^j - c_k^j)$, represent the center (cen) and half range (hr) values of $\tilde{\mathbf{x}}(k)$. Let the consequent parameters of the i-th rule (the parameters of the i-th local affine model) be represented as $\tilde{\mathbf{a}}_i = [\tilde{a}_0^i, \tilde{a}_1^i, \ldots, \tilde{a}_{r_c}^i]^\top$, where $\tilde{a}_j^i = [\gamma_i^j, \delta_i^j] \in \mathfrak{I}_{\gamma,\delta} = \{[\gamma,\delta] : \gamma, \delta \in \mathbb{R}, \gamma \leq \delta\}$. Moreover, assume $\mathbf{a}_i^{\text{cen}} = [a_0^{i,\text{cen}}, a_1^{i,\text{cen}}, \ldots, a_{r_c}^{i,\text{cen}}]^\top$, where $a_j^{i,\text{cen}} = 0.5(\gamma_i^j + \delta_i^j)$ and $\mathbf{a}_i^{\text{hr}} = [a_0^{i,\text{hr}}, a_1^{i,\text{hr}}, \ldots, a_{r_c}^{i,\text{hr}}]^\top$, where $a_j^{i,\text{hr}} = 0.5(\delta_i^j - \gamma_i^j)$, $j = 0, 1, \ldots, r_c$ and $i = 1, \ldots, c$, represent the center (cen) and half range (hr) values of $\tilde{\mathbf{a}}_i$. The vectors lumping all the center and half range consequent parameters are given by $\mathbf{a}^{\text{cen}} = [(\mathbf{a}_1^{\text{cen}})^\top, \ldots, (\mathbf{a}_c^{\text{cen}})^\top]^\top \in \mathbb{R}^{(r_c+1)c}$, where $\mathbf{a}_i^{\text{cen}} \in \mathbb{R}^{(r_c+1)}$, $i = 1, \ldots, c$ and $\mathbf{a}^{\text{hr}} = [(\mathbf{a}_1^{\text{hr}})^\top, \ldots, (\mathbf{a}_c^{\text{hr}})^\top]^\top \in \mathbb{R}^{(r_c+1)c}$, where $\mathbf{a}_i^{\text{hr}} \in \mathbb{R}^{(r_c+1)}$, $i = 1, \ldots, c$. These vectors are estimated globally by using the Ordinary Least Squares (OLS) method (for the NARX model). Denote the diagonal matrix $M_i \in \mathbb{R}^{N \times N}$ which has the membership grade $\mu_i(\tilde{\mathbf{z}}(k))$ as its k-th diagonal element with $1 \leq i \leq c$ and $1 \leq k \leq N$. Define the matrices

$$X_e^{\text{cen}} := [X^{\text{cen}}, \mathbf{1}] \in \mathbb{R}^{N \times (r_c+1)}, \tag{6.14}$$

$$X_e^{\text{hr}} := [X^{\text{hr}}, \mathbf{1}] \in \mathbb{R}^{N \times (r_c+1)}, \tag{6.15}$$

where $\mathbf{1}$ is a unitary column vector in $\mathbb{R}^n$, X^{cen} and X^{hr} are the input matrices for the center and radius consequent part, respectively.

$$X^{\text{cen}} := [\mathbf{x}^{\text{cen}}(1), \ldots, \mathbf{x}^{\text{cen}}(N)]^\top \in \mathbb{R}^{N \times r_c}, \tag{6.16}$$

$$X^{\text{hr}} := [\mathbf{x}^{\text{hr}}(1), \ldots, \mathbf{x}^{\text{hr}}(N)]^\top \in \mathbb{R}^{N \times r_c}. \tag{6.17}$$

Moreover, define

$$X_E^{\text{cen}} := [M_1 X_e^{\text{cen}}, \ldots, M_c X_e^{\text{cen}}] \in \mathbb{R}^{N\times(r_c+1)c}, \tag{6.18}$$

$$X_E^{\text{hr}} := [M_1 X_e^{\text{hr}}, \ldots, M_c X_e^{\text{hr}}] \in \mathbb{R}^{N\times(r_c+1)c}. \tag{6.19}$$

The reference interval output is defined as $\tilde{\mathbf{y}} = [\tilde{y}(1), \ldots, \tilde{y}(N)]^\top$, with $\tilde{y}(k) = [\underline{y}, \bar{y}] \in \mathfrak{I} = \{[s,t] : s,t \in \mathbb{R}, s \le t\}$, $k = 1, \ldots, N$. The center and half range values of $\tilde{\mathbf{y}}(k)$ are defined in the same way: $\mathbf{y}^{\text{cen}} = [y^{\text{cen}}(1), \ldots, y^{\text{cen}}(N)]^\top$ with $y^{\text{cen}}(k) = 0.5(\underline{y}(k) + \bar{y}(k))$, and $\mathbf{y}^{\text{hr}} = [y^{\text{hr}}(1), \ldots, y^{\text{hr}}(N)]^\top$ with $y^{\text{hr}}(k) = (\bar{y}(k) - \underline{y}(k))$, $k = 1, \ldots, N$. The center and half range values of the parameters of the local models ($\tilde{a}$) $\mathbf{a}^{\text{cen}}$ and $\mathbf{a}^{\text{hr}}$ are calculated as

$$\mathbf{a}^{\text{cen}} = [(X_E^{\text{cen}})^\top X_E^{\text{cen}}]^{-1}(X_E^{\text{cen}})^\top \mathbf{y}^{\text{cen}}, \tag{6.20}$$

$$\mathbf{a}^{\text{hr}} = [(X_E^{\text{hr}})^\top X_E^{\text{hr}}]^{-1}(X_E^{\text{hr}})^\top \mathbf{y}^{\text{hr}}. \tag{6.21}$$

6.4 Estimation of the NOE Model from NARX Model

The model estimated in the previous step is the NARX model which is evaluated in the scheme shown in Fig. 6.3. The NARX model predicts one step ahead interval outputs

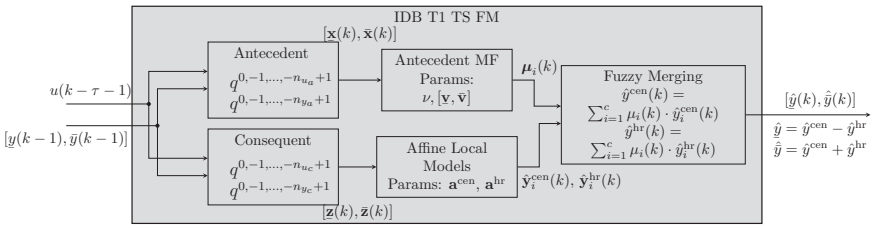


Figure 6.3: The evaluation scheme of the NARX model

($\tilde{\tilde{y}}(k)$) using the past inputs and past measured interval outputs. Having estimated the NARX model, its performance is evaluated for the NOE model scheme as illustrated in Fig. 6.4. As apparent from the figure, the NOE model uses the input signal and the past predictions to make the future predictions. The model estimated for the NARX case is usually not satisfactory for the NOE evaluation, in such a case, nonlinear optimization is used to estimate the model parameters for the NOE model as shown in Fig. 6.5. The parameters of

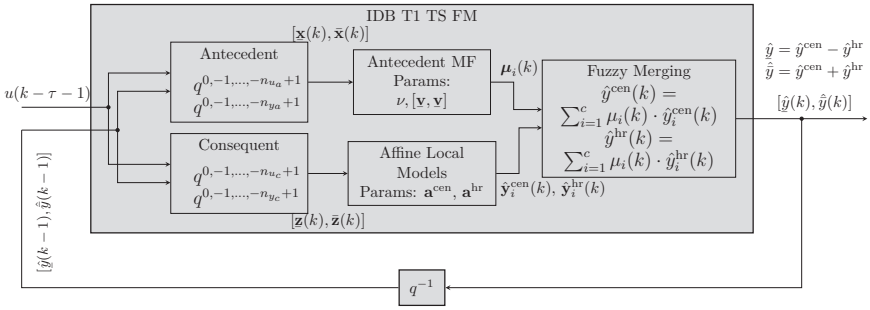


Figure 6.4: The evaluation scheme of the NOE model

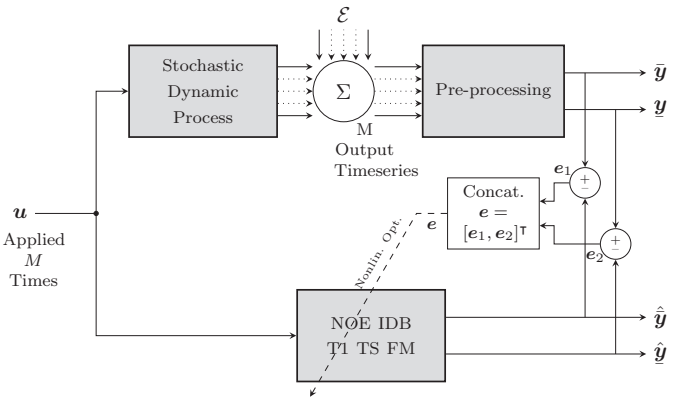


Figure 6.5: The estimation of the NOE model

the NARX model are used as the initial estimate. The reason why models are not trained for the NOE case by default is the resulting complex non-linear recursive optimization problem [38]. Furthermore, since the optimization results are highly dependent and sensitive on initial values, the parameters of the NARX model are passed to the nonlinear optimization as the best initial guess. The parameters to be optimized include cluster prototypes and local model parameters.

The center and half range of the cluster prototypes $\tilde{\mathbf{v}} = [\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_c]^\intercal$ are defined as $\mathbf{v}^{\text{cen}} = [(\mathbf{v}_1^{\text{cen}})^\intercal, \ldots, (\mathbf{v}_c^{\text{cen}}))^\intercal] \in \mathbb{R}^{cr_a}$, where $\mathbf{v}_i^{\text{cen}} \in \mathbb{R}^{r_a}$; and $\mathbf{v}^{\text{hr}} = [(\mathbf{v}_1^{\text{hr}})^\intercal, \ldots, (\mathbf{v}_c^{\text{hr}}))^\intercal] \in \mathbb{R}^{cr_a}$, where $\mathbf{v}_i^{\text{hr}} \in \mathbb{R}^{r_a}$; $i = 1, \ldots, c$. Denoting the lumped parameter vector of a NARX model

as $\boldsymbol{\theta}_{\text{NARX}} := [(\mathbf{v}^{\text{cen}})^\top, (\mathbf{v}^{\text{hr}})^\top, (\mathbf{a}^{\text{cen}})^\top, (\mathbf{a}^{\text{hr}})^\top]^\top$, with $\boldsymbol{\theta}_{\text{NARX}} \in \mathbb{R}^{2c(r_a+r_c+1)}$, the optimal set of parameters for the NOE model $\boldsymbol{\theta}_{\text{NOE}}$ is obtained by minimizing the mean quadratic prediction error of the NOE model for $\tilde{\mathbf{y}} = [\underline{\mathbf{y}}, \bar{\mathbf{y}}]$:

$$\boldsymbol{\theta}_{\text{NOE}} := \boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{k=1}^{N} (\bar{y}(k) - \hat{\bar{y}}_{\text{NOE}}(\boldsymbol{\theta}, k))^2 + (\underline{y}(k) - \hat{\underline{y}}_{\text{NOE}}(\boldsymbol{\theta}, k))^2. \tag{6.22}$$

In this research, the Matlab function `lsqnonlin` is used for solving the optimization problem in (6.22). The Levenberg-Marquardt method [55, 56] is selected as nonlinear optimization algorithm.

6.5 Case Studies

The results of IDB T1 TS FM in three case studies are provided in this section. For the general description of the test stands and the DOE see Chap. 2.

6.5.1 Academic Example

The input signal $\mathbf{u}$ of length $N = 2000$ applied to this system is shown in Chap. 2. The experiment is repeated $M = 100$ times which leads to a family of output time series given in Chap. 3. The extended Chebyshev's inequality with $\alpha = 0.25$ is applied to the M time series which transforms the output signal $N \times M$ manifold in the form of an interval, i.e. a $N \times 2$ matrix, where the first and second columns represent the lower and upper time series, respectively. The value of fuzziness parameter ($\nu = 1.3$), and the number of local models $c = 6$ is selected to give a good compromise between the model complexity and the approximation accuracy. It is demonstrated in Fig. 6.6, where the number of local models are plotted against the NMSE of the NOE model for the test data set. The antecedent and consequent variables are given by:

$$\tilde{\mathbf{z}}(k) = \tilde{\mathbf{x}}(k) = [u(k-1), \tilde{y}(k-1), \tilde{y}(k-2)]^\top$$

The Sequential Forward Selection (SFS) [27, 89] method was used for selecting these variables, in which each input is selected sequentially to optimize the total squared error. The

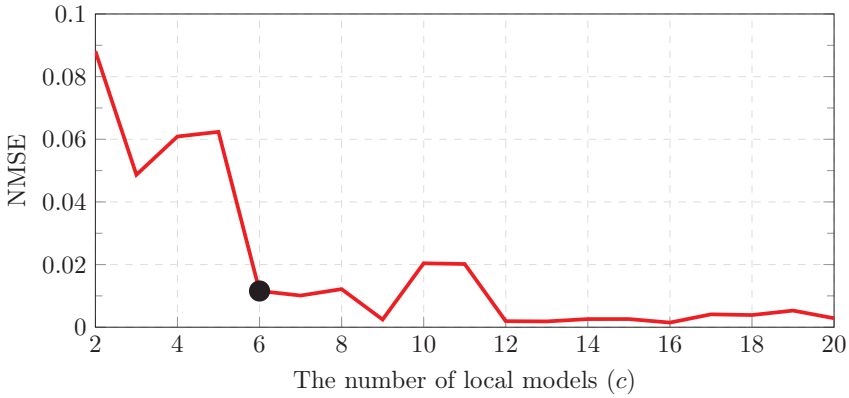


Figure 6.6: The number of local models (c) versus the NMSE of the NOE model for the test data set

clustering is performed in the input space. The first half of the data is used for identification, whereas the second half is used for validation.

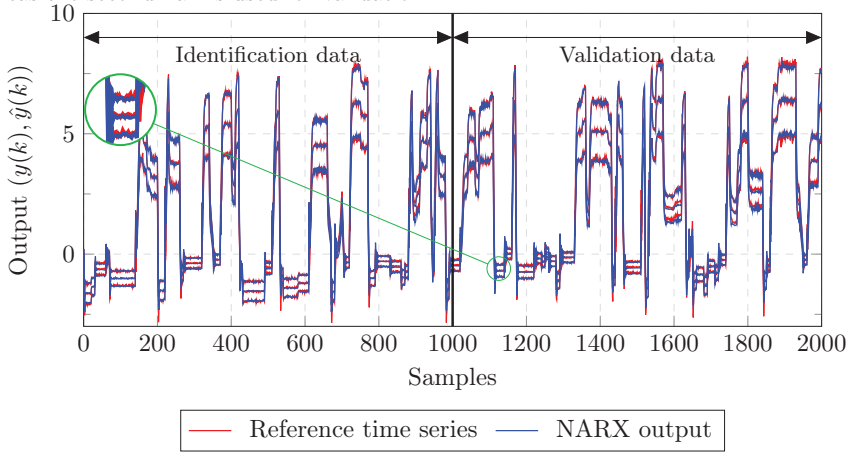


Figure 6.7: The output of the NARX model for the academic example

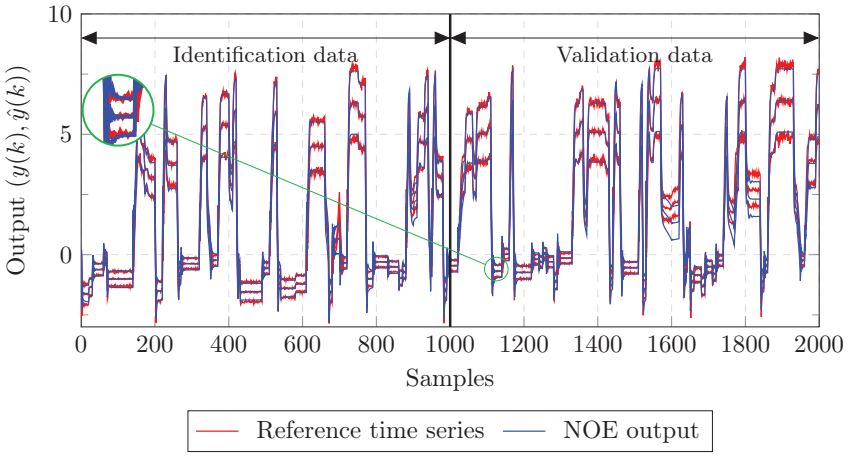


Figure 6.8: The output of the NOE model before optimization for the academic example

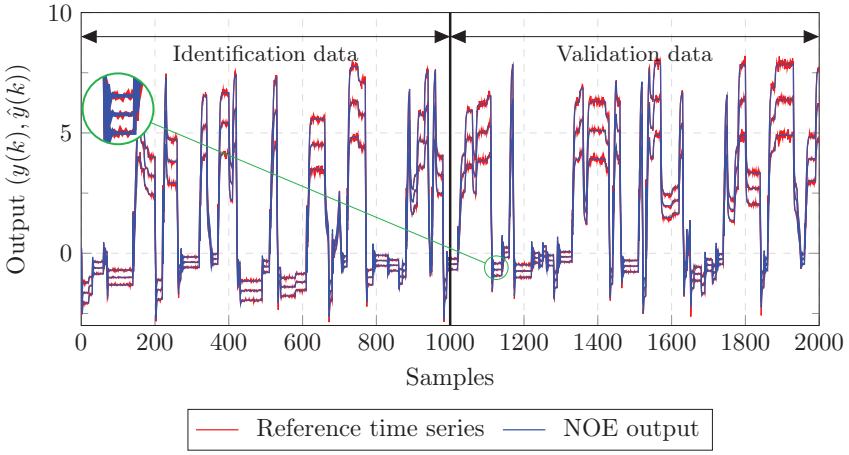


Figure 6.9: The output of the NOE model after optimization for the academic example

The results of the NARX model are illustrated in Fig. 6.7. The results of the NOE model before and after optimization are depicted in Fig. 6.8 and Fig. 6.9. Theses models are

assessed using the indices mentioned in Chap. 4, and the results are shown in Tab. 6.1. The error is defined as follows:

$$\mathbf{e}(k) = \mathbf{y}_e - \hat{\mathbf{y}}_e \tag{6.23}$$

where $\mathbf{y}_e = [\underline{\mathbf{y}}, \bar{\mathbf{y}}]^\top$ and $\hat{\mathbf{y}}_e = [\hat{\underline{\mathbf{y}}}, \hat{\bar{\mathbf{y}}}]^\top$.

Type of IDB T1 TS FLS	Identification Data			Validation Data		
	MaxAE	NMSE ($\times 10^{-3}$)	RMSE	MaxAE	NMSE ($\times 10^{-3}$)	RMSE
NARX	1.456	4.485	0.183	1.520	4.355	0.183
NOE_WO	2.344	11.254	0.289	2.250	14.025	0.328
NOE_OPT	0.410	0.607	0.067	0.455	0.863	0.081

Table 6.1.: Values of MaxAE, NMSE and RMSE for the NARX, NOE_WO (Without Optimization) and NOE_OPT (Optimized) models for the academic case study

6.5.2 Electro-mechanical Throttle Valve

The input signal $\mathbf{u}$ of length $N = 1000$ applied to this system is shown in Chap. 2. The sampling time is chosen to be $T_s = 10$ ms. The experiment is repeated $M = 80$ times. The value of $\alpha = 0.25$ is chosen for the extended Chebyshev's inequality. The value of the fuzziness parameter $\nu = 1.3$ is selected. The number of the local models $c = 8$ is selected to give a good compromise between the model complexity and the approximation accuracy as demonstrated in Fig. 6.10. The antecedent and consequent variables are given by:

$$\tilde{\mathbf{z}}(k) = \tilde{\mathbf{x}}(k) = [u(k-1), \tilde{y}(k-1), \tilde{y}(k-2)]^\top$$

It is a usual practice to have the same antecedent and consequent variables. However, since the Sequential Forward Selection (SFS) [27, 89] method is applied separately on both the antecedent and consequent variables, different antecedent and consequent variables were obtained. The clustering is performed in the input space. The first 900 data points are used for identification, whereas the remaining 100 data points are used for validation. The results of the NARX model are demonstrated in Fig. 6.11. The results of the NOE model before and after optimization are depicted in Fig. 6.12 and Fig. 6.13, respectively. The results of these models are shown in Tab. 6.2.

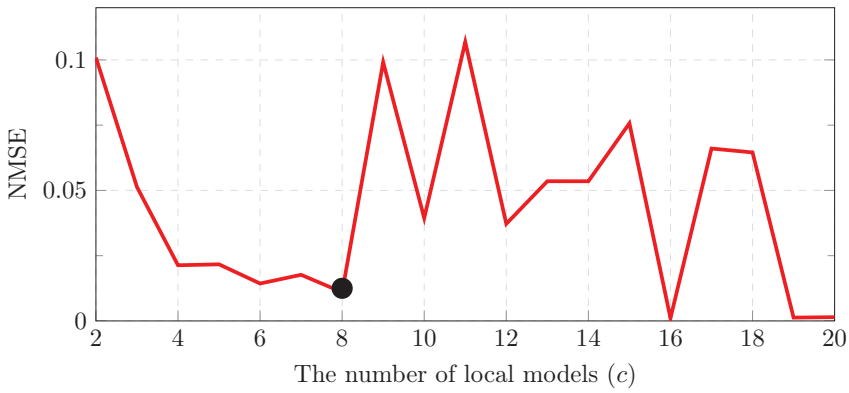


Figure 6.10: The number of local models (c) versus the NMSE of the NOE model for the test data set

Type of CDB T1 TS FLS	Identification Data			Validation Data		
	MaxAE in °	NMSE ($\times 10^{-3}$)	RMSE in °	MaxAE in °	NMSE ($\times 10^{-3}$)	RMSE in °
NARX	1.224	0.022	0.081	0.165	0.075	0.052
NOE_WO	5.431	12.008	1.911	2.006	21.239	0.868
NOE_OPT	1.660	1.028	0.559	1.328	7.018	0.499

Table 6.2.: Values of MaxAE, NMSE and RMSE for the NARX, NOE_WO (Without Optimization) and NOE_OPT (Optimized) models for the electro-mechanical throttle case study

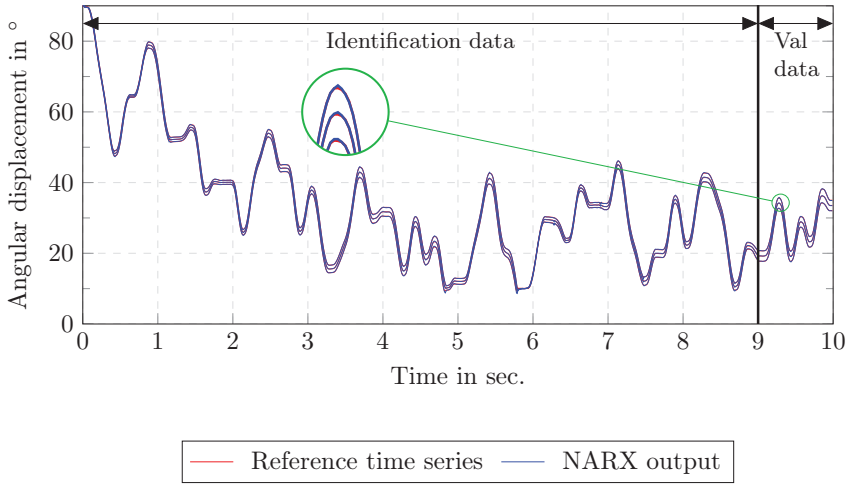


Figure 6.11: The output of the NARX model for the electro-mechanical throttle

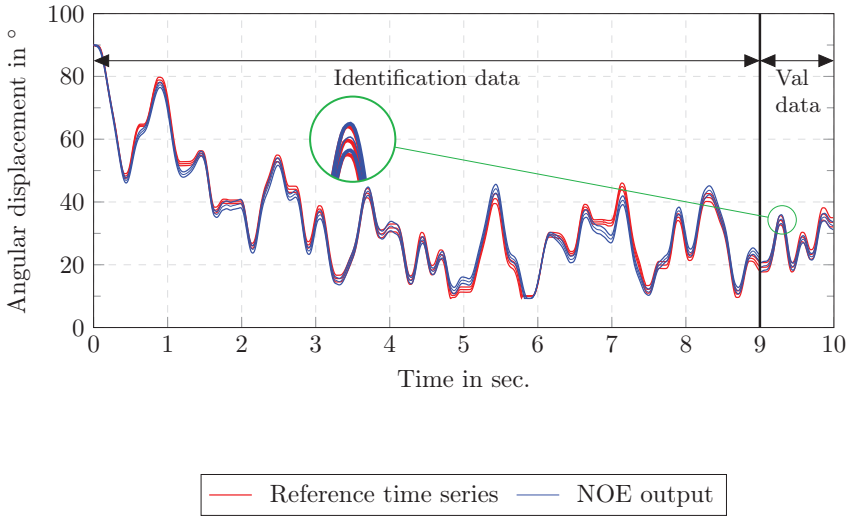


Figure 6.12: The output of the NOE model before optimization for the electro-mechanical throttle

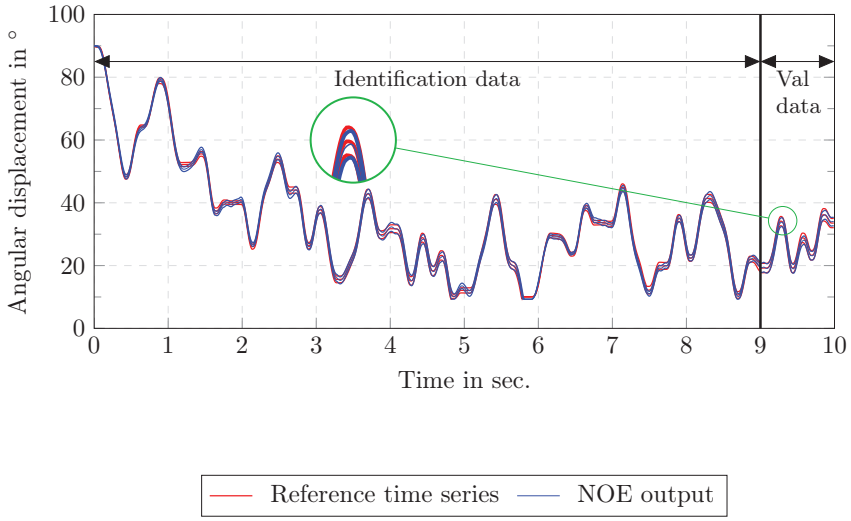


Figure 6.13: The output of the NOE model after optimization for the electro-mechanical throttle

6.5.3 Servo-Pneumatic Longitudinal Drive

The input signal $\mathbf{u}$ of length $N = 800$ applied to this system is shown in Chap. 2. The sampling time is chosen to be $T_s = 0.1$ s. The experiment is repeated $M = 100$ times which leads to a family of output time series given in Chap. 2. In this case study, the value of the fuzziness parameter $\nu = 1.1$ is selected. The number of the local models $c = 10$ is selected to give a good compromise between the model complexity and the approximation accuracy as demonstrated in Fig. 6.14. The antecedent and consequent variables is selected by the

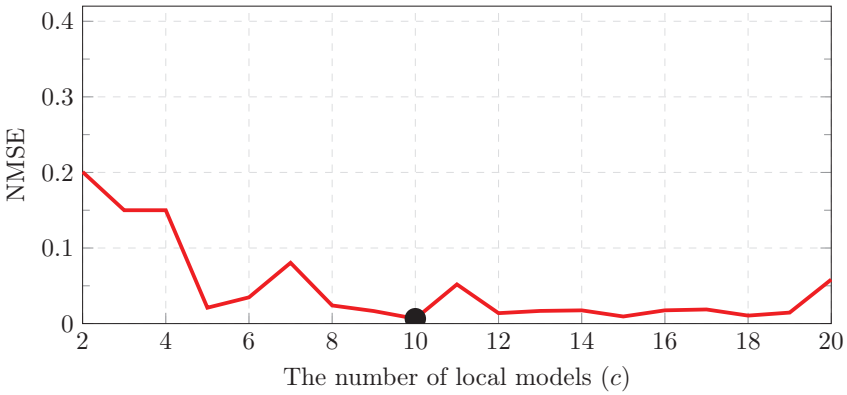


Figure 6.14: The number of local models (c) versus the NMSE of the NOE model for the test data set

Sequential Forward Selection (SFS) [27, 89] and given by:

$$\tilde{\mathbf{z}}(k) = [u(k-1), \tilde{y}(k-1), \tilde{y}(k-2)]^{\mathsf{T}}$$
$$\tilde{\mathbf{x}}(k) = [u(k-1), u(k-2), \tilde{y}(k-1), \tilde{y}(k-2), \tilde{y}(k-3)]^{\mathsf{T}}$$

The clustering is performed in the input space. The first 700 data points are used for identification, whereas the rest of 100 data points are used for validation. The results of the NARX model are demonstrated in Fig. 6.15. The results of the NOE model before and after optimization are depicted in Fig. 6.16 and Fig. 6.17. The modeling results are shown in Tab. 6.3.

Type of CDB T1 TS FLS	Identification Data			Validation Data		
	MaxAE in m	NMSE ($\times 10^{-3}$)	RMSE in m	MaxAE in m	NMSE ($\times 10^{-3}$)	RMSE in m
NARX	0.027	1.328	0.005	0.026	2.419	0.006
NOE_WO	0.068	34.270	0.025	0.087	136.654	0.047
NOE_OPT	0.040	6.057	0.010	0.059	23.002	0.019

Table 6.3.: Values of MaxAE, NMSE and RMSE for the NARX, NOE_WO (Without Optimization) and NOE_OPT (Optimized) models for the servo-pneumatic longitudinal drive case study

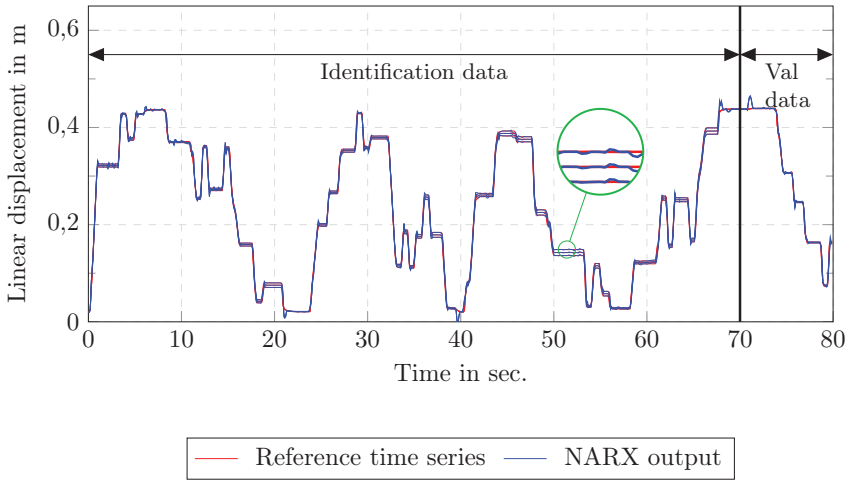


Figure 6.15: The output of the NARX model for the servo-pneumatic longitudinal drive

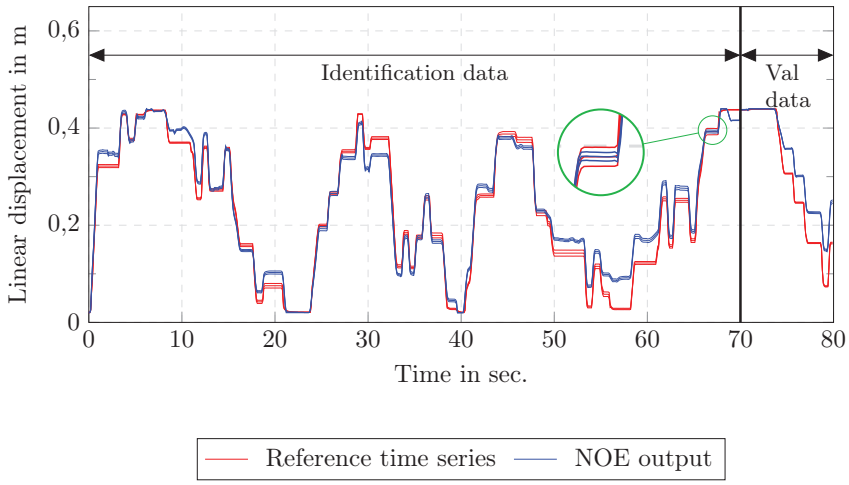


Figure 6.16: The output of the NOE model before optimization for the servo-pneumatic longitudinal drive

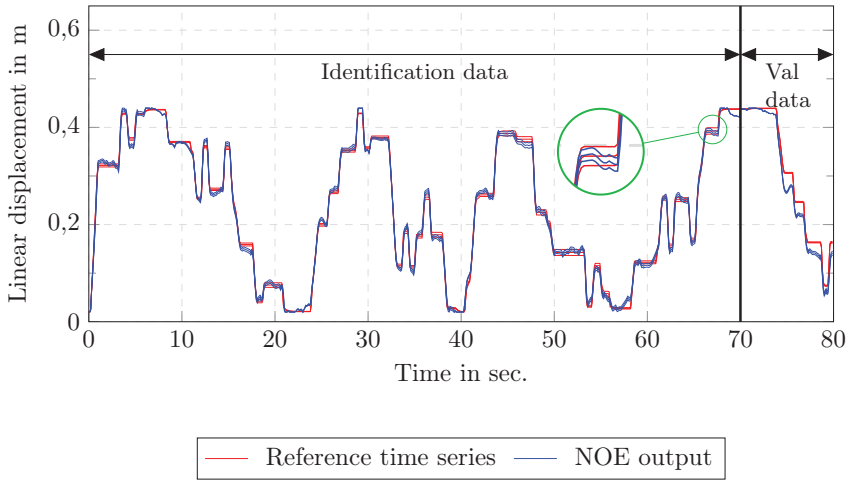


Figure 6.17: The output of the NOE model after optimization for the servo-pneumatic longitudinal drive

6.6 Summary and Discussion

The modeling of dynamic systems using IVD T1 TS FLS is described in this chapter. This model uses interval-valued data to estimate the T1 based multi-dimensional NOE TS model. The gist of the method lies in the repetition of the experiment. Using pre-processing the output time series are converted into lower and upper bound time series. The modeling results in the three case studies demonstrated that the presented modeling methodology is able to capture the lower bound, the upper bound and the mean time series. The modeling results can further be improved by designing an optimal input signal, increasing the signal length, tuning the parameters of the nonlinear optimization algorithm, and the selection of optimal model structure.

7

Conclusions and Outlook

This dissertation presents various approaches to build mathematical models using T1 and IT2 TS FM that can provide the mean response of an uncertain nonlinear dynamic system along with the spread around the mean as the measure of variability of the system output. The major contributions include the description of the system output as interval, rather a than crisp value, in order to capture the uncertainty in the observed data, without using an unrealistic assumption of normality, and to use this interval data to develop novel TS modelling frameworks, which are capable of handling the interval data. The methods of uncertainty modeling as well as TS fuzzy modeling using crisp and interval data have been presented and demonstrated for a nonlinear academic benchmark system, an electromechanical throttle and a servo-pneumatic longitudinal drive. By using the probabilistic analysis, it has been observed that the obtained PDFs are not 'nice' enough and too demanding to be handled by the existing probabilistic techniques. These PDFs are mostly non-normal, multimodal and often unsymmetric making it extremely difficult and computationally intractable to be used in the existing probabilistic modeling techniques. In order to come up with this conclusion, it was essential to first perform a probabilistic analysis, and thus to decide on the appropriate method. Moreover, the probabilistic analysis can be used in the future to improve the modeling even further, e.g. by incorporating fuzzy or histogram type of data in the fuzzy modelling. The presented approach in this thesis can use either max-min bounds, percentiles of data, or the extended Chebyshev inequality to obtain upper and lower bounds of the output time series. These bounds are directly used in the TS fuzzy model using the theory of symbolic interval-valued data. The developed model is referred to as the Interval-Data Based (IDB) Type-1 (T1) TS FM and its comparison with the other

possible alternative model description are presented in this research. The IDB T1 TS FM directly uses interval data in the modeling procedure by utilization the techniques of fuzzy clustering and regression on symbolic interval-valued data. This model is unique since it provides the estimates of the upper and lower bounds, whereas the mean response can be calculated by averaging them. Consequently, the obtained parameters are in the form of an interval. The results show that the presented approach is able to adequately model the stochastic effects due to the variability in the system's output in a unified efficient modelling framework. Apart from IDB T1 TS FM, the detailed modeling procedure using classical CDB T1 TS FM has been presented for all the case studies. Moreover, the conceptual sketches of IT2 modeling using crisp and interval data have been presented.

In the future, the modeling framework using crisp and interval data for IT2 TS FM can be investigated. Because of the superior uncertainty handling capability of type-2 FLS as compared to its counterpart type-1 FLS [42], the extension of type-1 FLS to T2 FLS for the developed modelling cases will be quite interesting. Since such models should be able to handle both crisp and interval data, and should have memberships expressed by intervals, special fuzzy clustering methods will be utilized. The effect and interpretation of uncertain cluster prototypes on modeling results will be explored. It will be interesting to observe the type-reduced fuzzy set of system output and to study its interpretation. In addition, the method will require special treatment for regression analysis in order to estimate consequent parameters. Moreover, some other techniques for obtaining less-conservative and robust upper and lower bounds from systems can be determined. In addition to developing models only for the mean values and the envelopes, the possibility of using histograms at each time instant directly in the modeling procedure using the theory of Symbolic Data Analysis can be explored. It may also be interesting how the PDFs defined by GMMs / KDE can be used instant-wise in modeling. Lastly, the input-output model description can be changed to a state space description to analyze the uncertainty in terms of state matrices, which can be beneficial for robust control. This concludes the outlook and some directions for future research.

A

Appendices

A.1 Proof of the Chebyshev's inequality

In order to prove the Chebyshev's inequality, the prove of Markov's inequality will be given first.

The Markov's inequality is given by:

$$P_{\mathsf{r}}(\mathsf{r} \geq a) \leq \frac{\mathbb{E}[\mathsf{r}]}{a} \tag{A.1}$$

Where P_{r} is the probability of the random variable r and $a \in \mathbb{R}^{>0}$. Consider the indicator function $a \cdot 1_{\mathsf{r} \geq \mathsf{a}}$ with the amplitude a given by:

$$a \cdot 1_{\mathsf{r} \geq \mathsf{a}} = \begin{cases} a & \textbf{if } \mathsf{r} \geq a \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

This function is illustrated in Fig. A.1. From the indicator function it is apparent that:

$$a \cdot 1_{\mathsf{r} \geq \mathsf{a}} \leq \mathsf{r} \tag{A.3}$$

Taking the expected value $\mathbb{E}[\cdot]$ on both the sides of (A.3)

$$\mathbb{E}[a \cdot 1_{\mathsf{r} \geq \mathsf{a}}] \leq \mathbb{E}[\mathsf{r}] \tag{A.4}$$

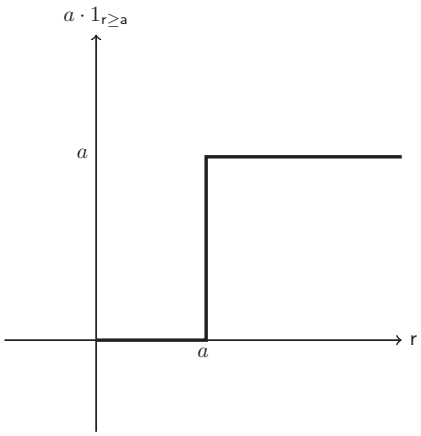


Figure A.1: The indicator function

Since $\mathbb{E}\left[a \cdot 1_{\mathsf{r}\geq\mathsf{a}}\right] = a \cdot \mathbb{E}\left[1_{\mathsf{r}\geq\mathsf{a}}\right]$ and $\mathbb{E}\left[1_{\mathsf{r}\geq\mathsf{a}}\right]$ is the definition of $P_{\mathsf{r}}(\mathsf{r} \geq a)$, the equation (A.4) cab be written as:

$$P_{\mathsf{r}}(\mathsf{r} \geq a) \leq \frac{\mathbb{E}\left[\mathsf{r}\right]}{a} \tag{A.5}$$

which concludes the proof of the Markov's inequality. Now the Chebyshev's inequality can be derived from (A.5). Replacing $\mathsf{r} \to |\mathsf{r} - \mu_x|$ in (A.5) yields the following:

$$P_{\mathsf{r}}(|\mathsf{r} - \mu_x| \geq a) \leq \frac{\mathbb{E}\left[|\mathsf{r} - \mu_x|\right]}{a} \tag{A.6}$$

The (A.6) is actually identical to the following equation:

$$P_{\mathsf{r}}((\mathsf{r} - \mu_x)^2 \geq a^2) \leq \frac{\mathbb{E}\left[(\mathsf{r} - \mu_x)^2\right]}{a^2} \tag{A.7}$$

Finally, combining (A.6) and (A.7) leads to:

$$P_{\mathsf{r}}(|\mathsf{r} - \mu_x| \geq a) \leq \frac{\mathbb{E}\left[(\mathsf{r} - \mu_x)^2\right]}{a^2} \tag{A.8}$$

By definition, $\mathbb{E}\left[(\mathsf{r} - \mu_x)^2\right] = \sigma_x^2$. Putting this value in (A.8) and replacing $a = k{\cdot}\sigma_x$ concludes the proof of the Chebyshev's inequality:

$$P_{\mathsf{r}}(|\mathsf{r} - \mu_x| \geq k \cdot \sigma_x) \leq \frac{1}{k^2} \tag{A.9}$$

A.2 Proof of the coverage factor for the extended Chebyshev's inequality

The extended Chebyshev's inequality is given by:

$$P(|y_k - m_{y_k}| \geq k_{\mathrm{cf}}\ s_{y_k}) \leq \frac{1}{\sqrt{M(M+1)}}\left(\frac{M-1}{k_{\mathrm{cf}}^2}+1\right) \tag{A.10}$$

Equating the right hand side with the significance level α provides the equation of the coverage factor as follows:

$$\frac{1}{\sqrt{M(M+1)}}\left(\frac{M-1}{k_{\mathrm{cf}}^2}+1\right) = \alpha \tag{A.11}$$

$$\frac{M-1}{k_{\mathrm{cf}}^2} = \alpha\sqrt{M(M+1)} - 1 \tag{A.12}$$

$$k_{\mathrm{cf}} = \sqrt{\frac{M-1}{\alpha\sqrt{M(M+1)}-1}} \tag{A.13}$$

Published Articles

S. ZAIDI, A. KROLL: NOE TS Fuzzy Modelling of Nonlinear Dynamic Systems with Uncertainties using Symbolic Interval-valued data. *Applied Soft Computing*, vol. 57, pp. 353–362, Aug. 2017.

S. ZAIDI, A. KROLL: On Interval-valued-data Type-1 Takagi-Sugeno fuzzy systems for uncertain nonlinear dynamic system identification. *Automatisierungstechnik*, vol. 64 (6), pp. 418–427, Jun. 2016.

S. ZAIDI, A. KROLL: A Novel Approach to T-S Fuzzy Modeling of Nonlinear Dynamic Systems with Uncertainties using Symbolic Interval-Valued Outputs. In: *Proceedings of the 17th IFAC Symposium on System Identification (SysID)*, pp. 1196–1201, 19-21.10.2015, Beijing, China.

S. ZAIDI, A. KROLL: Electro-Mechanical Throttle as a Benchmark Problem for Nonlinear System Identification with Friction. In: *Proceedings of the 24th Workshop Computational Intelligence, GMA-FA 5.14, KIT Scientific Publishing*, pp. 173–186, 27-28.11.2014, Dortmund, Germany.

S. ZAIDI, A. KROLL: On Identifying Envelop Type Nonlinear Output Error Takagi-Sugeno Fuzzy Models for Dynamic Systems with Uncertainties. In: *Proceedings of the 19th IFAC World Congress*, pp. 3226–3231, 24-29.08.2014, Cape Town, South Africa.

S. ZAIDI, A. KROLL: On identifying nonlinear envelop type dynamical T-S fuzzy models for systems with uncertainties: method and application to electro-mechanical throttle. In: *Proceedings of the 23rd Workshop Computational Intelligence, GMA-FA 5.14, KIT Scientific Publishing*, pp. 129–143, 05-06.12.2013, Dortmund, Germany.

S. ZAIDI, A. KROLL, H.-J. SOMMER: On Description and Identification of Uncertainties in System Modeling with Fuzzy Logic. In: *Proceedings of the 22nd Workshop Computational Intelligence, GMA-FA 5.14, KIT Scientific Publishing*, pp. 179–199, 06-07.12.2012, Dortmund, Germany.

Bibliography

[1] F. de A.T. de Carvalho. “Fuzzy c-means clustering methods for symbolic interval data”. *Pattern Recognition Letters* 28.4 (2007), pp. 423–437.

[2] R. H. Abiyev and O. Kaynak. “Identification and control of time-varying plants using type-2 fuzzy neural system”. In: *Proceedings of the 2009 International Joint Conference on Neural Networks*. Atlanta, GA, June 2009, pp. 13–19.

[3] J. Ackermann. *Robust Control*. Berlin: Springer-Verlag, 1993.

[4] H. Akaike. “Information Theory and an Extension of the Maximum Likelihood Principle”. In: *Selected Papers of Hirotugu Akaike*. Ed. by E. Parzen, K. Tanabe, and G. Kitagawa. New York: Springer New York, 1998, pp. 199–213.

[5] T. W. Anderson and D. A. Darling. “Asymptotic theory of certain “goodness-of-fit” criteria based on stochastic processes”. *The Annals of Mathematical Statistics* 23(2) (June 1952), pp. 193–212.

[6] S. Arlot and A. Celisse. “A survey of cross-validation procedures for model selection”. *Statistics Surveys* 4 (1984), pp. 40–79.

[7] R. Babuka. *Fuzzy Modeling for Control*. Boston: Kluwer Academic Publishers, 1998.

[8] N. Baklouti and A. M. Alimi. “Motion Planning in Dynamic and Unknown Environment Using an Interval Type-2 TSK Fuzzy Logic Controller”. In: *Proceedings of the IEEE International Fuzzy Systems Conference*. London, July 2007, pp. 1–6.

[9] T. Bernd, M. Kleutges, and A. Kroll. “Nonlinear Black Box Modelling – Fuzzy Networks versus Neural Networks”. *Neural Computing & Applications* 8(2) (1999), pp. 151–162.

[10] T. Bernd, A. Kroll, and H. Schwarz. “LS-optimal fuzzy modelling and its application to pneumatic drives”. In: *Proceedings of the European Control Conf. (ECC)*. 1997, pp. 1272–1277.

[11] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981.

[12] L. Billard and E. Diday. *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. John Wiley & Sons, 2012.

[13] L. Billard and E. Diday. "From the Statistics of Data to the Statistics of Knowledge". *Journal of the American Statistical Association* 98.462 (2003), pp. 470–487.

[14] S. A. Billings. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. West Sussex, United Kingdom: John Wiley & Sons, 2013.

[15] H. H. Bock and E. Diday. *Analysis of Symbolic Data. In: Exploratory Methods for Extracting Statistical Information from Complex Data*. Berlin Heidelberg: Springer-Verlag, 2000.

[16] H. Bustince, E. Barrenechea, M. Pagola, and R. Orduna. "Construction of Interval Type 2 Fuzzy Images to Represent Images in Grayscale. False Edges". In: *Proceedings of the IEEE International Fuzzy Systems Conference*. London, July 2007, pp. 1–6.

[17] I. M. Chakravarti, R. G. Laha, and J. Roy. *Handbook of Methods of Applied Statistics*. New York: John Wiley & Sons, 1967.

[18] P. L. Chebyshev. "Des valeurs moyennes". *J. Math. Pures Appl.* 12(2) (1867), pp. 177–184.

[19] A. Colubi, C. Fernandez-Garcia, and M. A. Gil. "Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data". *IEEE Transactions on Fuzzy Systems* 10(3) (June 2002), pp. 384–390.

[20] M. Cruz, G. Peters, and P. Shevchenko. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. Wiley Handbooks in Financial Engineering and Econometrics. Wiley, 2015.

[21] J. Daafouz et al. "Switched and piecewise affine systems". In: *Handbook of Hybrid Systems Control: Theory, Tools, Applications*. London: Cambridge University Press, 2009, pp. 89–137.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1) (1977), pp. 1–38.

[23] J. S. Denker. *Constructing Random Numbers with an Arbitrary Distribution*. Last access on August 08, 2016. URL: `http://www.av8n.com/physics/arbitrary-probability.htm`.

[24] T. Dereli, A. Baykasoglu, K. Altun, A. Durmusoglu, and I. B. Türksen. "Industrial applications of type-2 fuzzy sets and systems: A concise review". *Computers in Industry* 62(2) (2011), pp. 125–137.

[25] P. Diamond and H. Tanaka. “Fuzzy Regression Analysis”. In: *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*. Ed. by R. Sowiski. Boston, MA: Springer US, 1998, pp. 349–387.

[26] S. Duym and J. Shoukens. “Design of excitation signals for the restoring force surface method”. *Mechanical Systems and Signal Processing* 9(2) (1995), pp. 139–158.

[27] X. Fang and L. Wang. “Feature Selection Based on Fisher Criterion and Sequential Forward Selection for Intrusion Detection”. *Revista de la Facultad de Ingenierìa U.C.V* 32(1) (2017), pp. 498–503.

[28] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. “A clustering technique for the identification of piecewise affine systems”. *Automatica* 39(2) (Feb. 2003), pp. 205–217.

[29] GMA-Benchmarking. `http://www.rst.e-technik.tu-dortmund.de/cms/de/Veranstaltungen/GMA-Fachausschuss/Benchmark/index.html`. Last visit August 31, 2018. 2014.

[30] H. Hagras. “A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots”. *IEEE Transactions on Fuzzy Systems* 12(4) (2004), pp. 524–539.

[31] N.-B. Heidenreich, A. Schindler, and S. Sperlich. “Bandwidth selection for kernel density estimation: a review of fully automatic selectors”. *AStA Advances in Statistical Analysis* 97(4).4 (2013), pp. 403–433.

[32] R. V. Hogg, E. Tanis, and D. Zimmerman. *Probability and Statistical Inference*. Essex: Pearson Education Limited, 2013.

[33] C. M. Hurvich and C. L. Tsai. “Regression and time series model selection in small samples”. *Biometrika* 76(2) (1989), pp. 297–307.

[34] C. Hwang, D. H. Hong, and K. H. Seok. “Support vector interval regression machine for crisp input and output data”. *Fuzzy Sets and Systems* 157(8) (2006), pp. 1114–1125.

[35] J. T. G. Hwang and A. A. Ding. “Prediction Intervals for Artificial Neural Networks”. *Journal of the American Statistical Association* 92(438) (1997).

[36] S. Jafarinezhad and M. Shahbazian. “System identification of a Non-Linear Continuous Stirred Tank Heater based on Type-2 Fuzzy System”. In: *Proceedings of the Chinese Control and Decision Conference (CCDC)*. Mianyang, May 2011, pp. 1869–1874.

[37] S. Jafarzadeh, M. S. Fadali, M. Etezadi-Amoli, and A. Nafeh. “Type 1 and type 2 fuzzy TSK modeling of solar radiation for PV power generation”. In: *Proceedings of the North American Power Symposium (NAPS), 2010*. Arlington, TX, Sept. 2010, pp. 1–5.

[38] M. Jelali and A. Kroll. *Hydraulic Servo-systems: Modelling, Identification and Control*. London: Springer-Verlag, 2002.

[39] Z. Jianhua, C. Hongjie, and W. Rubin. "Nonlinear dynamical system identification based on evolutionary interval type-2 TSK fuzzy systems". In: *Proceedings of the Chinese Control and Decision Conference (CCDC)*. Qingdao, May 2015, pp. 2871–2876.

[40] A. John, M. Sugumaran, and R. S. Rajesh. "Indexing And Query Processing Techniques In Spatio-Temporal Data". *ICTACT Journal on Soft Computing* 6(3) (2016), pp. 1198–1217.

[41] A. Kabán. "Non-parametric detection of meaningless distances in high dimensional data". *Statistics and Computing* 22(2) (2012), pp. 375–385.

[42] N. N. Karnik and J. M. Mendel. "Applications of type-2 fuzzy logic systems: handling the uncertainty associated with surveys". In: *Proceedings of the IEEE International Fuzzy Systems Conference Proceedings (FUZZ-IEEE)*. Vol. 3. Seoul, Korea, Aug. 1999, pp. 1546–1551.

[43] N. N. Karnik, J. M. Mendel, and Q. L. Liang. "Type-2 fuzzy logic system". *IEEE Transactions on Fuzzy Systems* 7(6) (Dec. 1999), pp. 643–658.

[44] N. N. Karnik and J. M. Mendel. "Applications of type-2 fuzzy logic systems to forecasting of time-series". *Information Sciences* 120 (1–4) (1999), pp. 89–111.

[45] E. Kayacan, E. Kayacan, and M. A. Khanesar. "Identification of Nonlinear Dynamic Systems Using Type-2 Fuzzy Neural Networks – A Novel Learning Algorithm and a Comparative Study". *IEEE Transactions on Industrial Electronics* 62(3) (Mar. 2015), pp. 1716–1724.

[46] A. Khosravi and S. Nahavandi. "An interval type-2 fuzzy logic system-based method for prediction interval construction". *Applied Soft Computing* 24 (2014), pp. 222–231.

[47] A. Khosravi, S. Nahavandi, and D. Creighton. "Prediction Interval Construction and Optimization for Adaptive Neurofuzzy Inference Systems". *IEEE Transactions on Fuzzy Systems* 19(5) (Oct. 2011), pp. 983–988.

[48] 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* 22(3) (Mar. 2011), pp. 337–346.

[49] A. Khosravi, S. Nahavandi, D. Creighton, and R. Naghavizadeh. "Prediction interval construction using interval type-2 Fuzzy Logic systems". In: *Procedings of the 2012 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Brisbane, QLD, 2012, pp. 1–7.

[50] H. S. Konijn. “Distribution-Free and Other Prediction Intervals”. *The American Statistician* 41(1) (1987), pp. 11–15.

[51] A. Kroll. *Grey-box models: Concepts and application. In: New Frontiers in Computational Intelligence and its Applications*. Amsterdam: IOS Press. Amsterdam: IOS Press, 2000, pp. 42–51.

[52] A. Kroll. “Identification of functional fuzzy models using multidimensional reference fuzzy sets”. *Fuzzy Sets and Systems* 80(2) (June 1996), pp. 149–158.

[53] A. Kroll and H. Schulte. “Benchmark problems for nonlinear system identification and control using Soft Computing methods: Need and overview”. *Applied Soft Computing* 25 (2014), pp. 496–513.

[54] M. Lebbal, H. Chafouk, G. Hoblos, and D. Lefebvre. “Modeling and identification of non-linear systems by a multimodel approach: :application to a throttle valve”. 3(1) (2006), pp. 67–87.

[55] K. Levenberg. “A Method for the Solution of Certain Non-Linear Problems in Least Squares”. *Quarterly of Applied Mathematics* 2 (1944), pp. 164–168.

[56] K. Levenberg. “An Algorithm for Least-Squares Estimation of Nonlinear Parameters”. *SIAM Journal on Applied Mathematics* 11(2) (1963), pp. 431–441.

[57] H.-X. Li, X. Duan, and Z. Liu. “Three-dimensional fuzzy logic system for process modeling and control”. *Journal of Control Theory and Applications* 8(3) (2010), pp. 280–285.

[58] P. Liang and F. Song. “What does a probabilistic interpretation of fuzzy sets mean?” *IEEE Transactions on Fuzzy Systems* 4(2) (May 1996), pp. 200–205.

[59] Q. Liang and J. Mendel. “Interval type-2 fuzzy logic systems: theory and design”. *IEEE Transactions on Fuzzy Systems* 8(5) (2000), pp. 535–550.

[60] Q. Liang and J. M. Mendel. “An introduction to type-2 TSK fuzzy logic systems”. In: *Proceedings of the IEEE International Fuzzy Systems Conference Proceedings (FUZZ-IEEE)*. Vol. 3. Seoul, South Korea, 1999, pp. 1534–1539.

[61] Q. Liang and J. M. Mendel. “Decision feedback equalizer for nonlinear time-varying channels using type-2 fuzzy adaptive filters”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Vol. 2. San Antonio, TX, 2000, pp. 883–888.

[62] Q. Liang and J. M. Mendel. “Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters”. *IEEE Transactions on Fuzzy Systems* 8(5) (Oct. 2000), pp. 551–563.

[63] H. Lilliefors. “On the kolmogorovsmirnov test for normality with mean and variance unknown”. *Journal of the American Statistical Association* 62 (June 1967), pp. 399–402.

[64] E. d. A. Lima Neto and F. d. A. de Carvalho. “Centre and Range method for fitting a linear regression model to symbolic interval data”. *Computational Statistics & Data Analysis* 52(3) (2008), pp. 1500–1515.

[65] O. Linda and M. Manic. “On the accuracy of input-output uncertainty modeling with interval Type-2 Fuzzy Logic Systems”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Brisbane, QLD, June 2012, pp. 1–7.

[66] Z. Liu and H. X. Li. “A probabilistic fuzzy logic system for modeling and control”. *IEEE Transactions on Fuzzy Systems* 13(6) (2005), pp. 848–859.

[67] Z. Liu and H.-X. Li. “Probabilistic fuzzy logic system: A tool to process stochastic and imprecise information”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Jeju Island, 2009, pp. 848–853.

[68] E. Mamdani and S. Assilian. “An experiment in linguistic synthesis with a fuzzy logic controller”. *International Journal of Man-Machine Studies* 7(1) (1975), pp. 1–13.

[69] G. McLachlan and D. Peel. *Finite Mixture Models*. Hoboken, NJ: John Wiley & Sons, Inc., 2000.

[70] A. H. Meghdadi and M. R. Akbarzadeh-T. “Probabilistic fuzzy logic and probabilistic fuzzy systems”. In: *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*. Vol. 3. 2001, pp. 1127–1130.

[71] J. M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper-Saddle River, NJ: Prentice Hall, 2001.

[72] J. M. Mendel. “Type-2 fuzzy sets and systems: an overview”. *IEEE Computational Intelligence Magazine* 2.1 (Feb. 2007), pp. 20–29.

[73] J. M. Mendel and R. I. B. John. “Type-2 fuzzy sets made simple”. *IEEE Transactions on Fuzzy Systems* 10(2) (Apr. 2002), pp. 117–127.

[74] J. Mendel and F. Liu. “On new quasi-type-2 fuzzy logic systems”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. 2008, pp. 354–360.

[75] J. Mendel, F. Liu, and D. Zhai. “α-Plane Representation for Type-2 Fuzzy Sets: Theory and Applications”. *IEEE Transactions on Fuzzy Systems* 17(5) (2009), pp. 1189–1207.

[76] J. M. Mendel. “Uncertainty, fuzzy logic, and signal processing”. *Signal Processing* 80(6) (2000), pp. 913–933.

[77] S. Naim and H. Hagras. “A hybrid approach for Multi-Criteria Group Decision Making based on interval type-2 fuzzy logic and Intuitionistic Fuzzy evaluation”. In: *Proceedings of the 2012 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Brisbane, QLD, June 2012, pp. 1–8.

[78] S. Naim and H. Hagras. “A general type-2 fuzzy logic based approach for Multi-Criteria Group Decision Making”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ)*. Hyderabad, July 2013, pp. 1–8.

[79] K. Narendra and K. Parthasarathy. “Adaptive identification and control of dynamical systems using neural networks”. In: *Proceedings of the 28th IEEE Conference on Decision and Control*. Tampa, FL, 1989, pp. 1737–1738.

[80] K. Narendra and K. Parthasarathy. “Identification and control of dynamical systems using neural networks”. *IEEE Transactions on Neural Networks* 1(1) (Mar. 1990), pp. 4–27.

[81] O. Nelles. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin Heidelberg: Springer-Verlag, 2001.

[82] T. Nguyen, A. Khosravi, S. Nahavandi, and D. Creighton. “Neural network and interval type-2 fuzzy system for stock price forecasting”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ)*. Hyderabad, July 2013, pp. 1–8.

[83] M. Nie and W. W. Tan. “Modeling capability of type-1 fuzzy set and interval type-2 fuzzy set”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Brisbane, QLD, June 2012, pp. 1–8.

[84] D. A. Nix and A. S. Weigend. “Estimating the mean and variance of the target probability distribution”. In: *Proceedings of the IEEE World Congress on Computational Intelligence*. Vol. 1. Orlando, FL, June 1994, pp. 55–60.

[85] M. Norgaard, O. Ravn, N. K. Poulsen, and L. K. Hansen. *Neural networks for modelling and control of dynamic systems*. London: Springer, 2003.

[86] S. Nurmaini and S. Z. M. Hashim. “Motion planning in unknown environment using an interval fuzzy type-2 and neural network classifier”. In: *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*. Hong Kong, 2009, pp. 50–55.

[87] H. Olsson, K. Åström, C. C. de Wit, M. Gäfvert, and P. Lischinsky. “Friction Models and Friction Compensation”. *European Journal of Control* 4(3) (1998), pp. 176–195.

[88] P. M. Oscar Castillo. *Type-2 Fuzzy Logic: Theory and Applications*. Berlin: Springer-Verlag, 2008.

[89] S. Pal and P. Mitra. *Pattern Recognition Algorithms for Data Mining*. Chapman & Hall/CRC Computer Science & Data Analysis. London: CRC Press, 2004.

[90] M. M. Pelagatti. *Time Series Modelling with Unobserved Components*. Boca Raton, FL: CRC Press, 2016.

[91] N. M. Razali and Y. B. Wah. "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests". *Journal of Statistical Modeling and Analytics* 2(1) (2011), pp. 21–33.

[92] R. A. Redner and H. F. Walker. "Mixture Densities, Maximum Likelihood and the EM Algorithm". *SIAM Review* 26(2) (1984), pp. 195–239.

[93] D. Rees. "Automatic testing of dynamic systems using multifrequency signals and discrete Fourier transform". In: *Proceedings of the New Developments in Automatic Testing*. Brighton, UK, 1977, pp. 24–27.

[94] Z. Ren, A. Kroll, M. Sofsky, and F. Laubenstein. "On methods for automated modeling of dynamic systems with friction and their application to electro-mechanical throttles". In: *Proceedings of the 49th IEEE Conference on Decision and Control*. Atlanta, GA, USA, Dec. 2010, pp. 7637–7642.

[95] Z. Ren, A. Kroll, M. Sofsky, and F. Laubenstein. "On identification of piecewise-affine models for systems with friction and its application to electro-mechanical throttles". In: *Proceedings of the 16th IFAC Symposium on System Identification, SysID*. Brussel, Belgium, Dec. 2012, pp. 1395–1400.

[96] Z. Ren. "Zur Identifikation mechatronischer Stellglieder mit Reibung bei Kraftfahrzeugen". Dissertation. Schriftenreihe Mess- und Regelungstechnik der Universität Kassel, 2012.

[97] Z. Ren, A. Kroll, M. Sofsky, and F. Laubenstein. "Zur physikalischen und datengetriebenen Modellbildung von Systemen mit Reibung: Methoden und Anwendung auf Kfz-Drosselklappen". *at - Automatisierungstechnik* 61(3) (2013), pp. 155–171.

[98] T. J. Ross, J. M. Booker, and W. J. Parkinson, eds. *Fuzzy Logic and Probability Applications: Bridging the Gap*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2002.

[99] J. Russell and R. Cohn. *Chebyshev's Inequality*. Book on Demand, 2012.

[100] J. G. Saw, M. C. K. Yang, and T. C. Mo. "Chebyshev Inequality with Estimated Mean and Variance". *The American Statistician* 38(2) (1984), pp. 130–132.

[101] R. Scattolini, C. Siviero, M. Mazzucco, S. Ricci, L. Poggio, and C. Rossi. "Modeling and identification of an electromechanical internal combustion engine throttle body." *The American Statistician* 5(9) (1997), pp. 1253–1259.

[102] M. R. Schröder. “Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (Corresp.)” *IEEE Transactions on Information Theory* 16(1) (Jan. 1970), pp. 85–89.

[103] S. S. Shapiro and M. B. Wilk. “An analysis of variance test for normality (complete samples)”. *Biometrika* 52(2-4) (1965), pp. 591–611.

[104] B. W. Silverman. *Density estimation for statistics and data analysis*. London: Chapman & Hall/CRC, 1986.

[105] I. Skrjanc, S. Blazic, and O. Agamennoni. “Identification of dynamical systems with a robust interval fuzzy model”. *Automatica* 41(2) (2005), pp. 327–332.

[106] I. Skrjanc, S. Blazic, and O. Agamennoni. “Interval Fuzzy Model Identification Using l_∞ -Norm”. *IEEE Transactions on Fuzzy Systems* 13(5) (2005).

[107] C. Sun and Z. Xu. “Extended T-S fuzzy model based on interval arithmetic and its application to interval nonlinear regression analysis”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Jeju Island, Korea, 2009, pp. 1773–1778.

[108] T. Takagi and M. Sugeno. “Fuzzy identification of systems and its applications to modeling and control”. *IEEE Transactions on Systems, Man and Cybernetics* 15(1) (Jan. 1985), pp. 116–132.

[109] H. Tanaka, S. Uejima, and K. Asai. “Linear Regression Analysis with Fuzzy Model”. *IEEE Transactions on Systems, Man, and Cybernetics* 12(6) (Nov. 1982), pp. 903–907.

[110] L. H. Ungar, R. D. D. Veaux, and E. Rosengarten. “Estimating Prediction Intervals for Artificial Neural Networks”. In: *Proceedings of the 9th Yale Workshop on Adaptive and Learning Systems*. 1996.

[111] M. Vasak, L. Mladenovic, and N. Peric. “Clustering-based identification of a piecewise affine electronic throttle model”. In: *Proceedings of the 31st Annual Conference of IEEE, IECON*. Nov. 2005, pp. 177–182.

[112] C. Wagner and H. Hagras. “Toward General Type-2 Fuzzy Logic Systems Based on zSlices”. *IEEE Transactions on Fuzzy Systems* 18(4) (2010), pp. 637–660.

[113] L. X. Wang. “Fuzzy systems are universal approximators”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems*. Mar. 1992, pp. 1163–1170.

[114] M. L. Wang and H. B. Shi. “Proceedings of the fuzzy-neuro network based on improved nearest-neighborhood clustering algorithm”. In: *The 23rd Chinese Control Conference*. 2004, pp. 293–297.

[115] M. Wang, N. Li, and S. Li. “Type-2 T-S fuzzy modeling for the dynamic systems with measurement noise”. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Hong Kong, China, June 2008, pp. 443–448.

[116] M. Witters and J. Swevers. “Black-box model identification for a continuously variable, electro-hydraulic semi-active damper”. *Mechanical Systems and Signal Processing* 24(1) (2010), pp. 4–18.

[117] D. Wu. “On the Fundamental Differences Between Interval Type-2 and Type-1 Fuzzy Logic Controllers”. *IEEE Transactions on Fuzzy Systems* 20(5) (2012), pp. 832–848.

[118] D. Wu and W. W. Tan. “Type-2 FLS Modeling Capability Analysis”. In: *Proceedings of the 14th IEEE International Conference on Fuzzy Systems (FUZZ ’05)*. Reno, NV, May 2005, pp. 242–247.

[119] K. C. Wu. “Fuzzy interval control of mobile robots”. *Computers & Electrical Engineering* 22(3).3 (1996), pp. 211–229.

[120] Z. Xu and C. Sun. “Interval T-S fuzzy model and its application to identification of nonlinear interval dynamic system based on interval data”. In: *Proceedings of the 48th IEEE Conference on Decision and Control/Chinese Control Conference*. Shanghai, China, 2009, pp. 4144–4149.

[121] M. Yang and N. Ahuja. “Gaussian mixture model for human skin color and its applications in image and video databases”. In: *Procedings of the SPIE*. San Jose, CA, Dec. 1999, pp. 458–466.

[122] H. Ying. “General SISO Takagi-Sugeno fuzzy systems with linear rule consequent are universal approximators”. *IEEE Transactions on Fuzzy Systems* 6(4) (Nov. 1998), pp. 582–587.

[123] L. A. Zadeh. “Outline of a New Approach to the Analysis of Complex Systems and Decision Processes”. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (1) (Jan. 1973), pp. 28–44.

[124] L. A. Zadeh. “The Concept of a Linguistic Variable and its Application to Approximate Reasoning-1”. *Information Sciences* (1985), pp. 199–249.

[125] L. A. Zadeh. “Discussion: Probability theory and fuzzy logic are complentary rather than competitive”. *Technometrics* 37(3) (1995), pp. 271–276.

[126] L. A. Zadeh. “Toward a generalized theory of uncertainty (GTU)–an outline”. *Information Sciences* 172(1–2) (2005), pp. 1–40.

[127] L. Zadeh. “Fuzzy sets”. *Information and Control* 8(3) (1965), pp. 338–353.

[128] S. Zaidi and A. Kroll. “On identifying nonlinear envelop type dynamical T-S fuzzy models for systems with uncertainties: method and application to electro-mechanical

throttle". In: *Proceedings of the 23. Workshop Computational Intelligence*. Dortmund: KIT Scientific Publishing, Dec. 2013, pp. 129–143.

[129] S. Zaidi and A. Kroll. "Electro-Mechanical Throttle as a Benchmark Problem for Nonlinear System Identification with Friction". In: *Proceedings of the 24. Workshop Computational Intelligence*. Dortmund: KIT Scientific Publishing, Nov. 2014, pp. 173–186.

[130] S. Zaidi and A. Kroll. "On Identifying Envelop Type Nonlinear Output Error Takagi-Sugeno Fuzzy Models for Dynamic Systems with Uncertainties". In: *Proceedings of the 19th IFAC World Congress*. Cape Town, South Africa, Aug. 2014, pp. 3226–3231.

[131] S. Zaidi and A. Kroll. "A Novel Approach to T-S Fuzzy Modeling of Nonlinear Dynamic Systems with Uncertainties using Symbolic Interval-Valued Outputs". In: *Proceedings of the 17th IFAC Symposium on System Identification (SysID)*. Beijing, China, Oct. 2015, pp. 1196–1201.

[132] S. Zaidi and A. Kroll. "NOE TS Fuzzy Modelling of Nonlinear Dynamic Systems with Uncertainties using Symbolic Interval-valued data". *Applied Soft Computing* (2016). submitted.

[133] S. Zaidi and A. Kroll. "On Interval-valued-data Type-1 Takagi-Sugeno fuzzy systems for uncertain nonlinear dynamic system identification". *at - Automatisierungstechnik* 64(6) (2016), pp. 418–427.

[134] S. Zaidi, A. Kroll, and H.-J. Sommer. "On Description and Identification of Uncertainties in System Modeling with Fuzzy Logic". In: *Proceedings of the 22. Workshop Computational Intelligence*. Dortmund: KIT Scientific Publishing, Dec. 2012, pp. 179–199.

[135] H. F. Zarandi, I. Türken, and T. Kasbi". "Type-2 fuzzy modeling for desulphurization of steel process". *Expert Systems with Applications* 32(1) (2007), pp. 157–171.

[136] M. Zarinbal, L. B. Turksen, M. H. F. Zarandi, and M. Izadid. "Interval type-2 fuzzy image processing expert system for diagnosing brain tumors". In: *Proceedings of the IEEE Conference on Norbert Wiener in the 21st Century (21CW)*. Boston, MA, June 2014, pp. 1–8.

[137] G. Zhang and H. X. Li. "An Efficient Configuration for Probabilistic Fuzzy Logic System". *IEEE Transactions on Fuzzy Systems* 20(5) (Oct. 2012), pp. 898–909.

Schriftenreihe Mess- und Regelungstechnik der Universität Kassel

Herausgegeben von / Edited by
Univ.-Prof. Dr.-Ing. Andreas Kroll, Universität Kassel

Band 1: Klassifikationsgestützte on-line Adaption eines robusten beobachter-basierten Fehlerdiagnoseansatzes für nichtlineare Systeme, Kassel 2011
Patrick Gerland

Band 2: Zur Identifikation mechatronischer Stellglieder mit Reibung bei Kraftfahrzeugen, Kassel 2012
Zhenxing Ren

Band 3: Sensordatenfusionsansätze in der Thermografie zur Verbesserung der Messergebnisse, Kassel 2014
Samuel Soldan

Band 4: Multi-Robot Task Allocation for Inspection Problems with Cooperative Tasks Using Hybrid Genetic Algorithms, Kassel 2014
Chun Liu

Band 5: Gasleckortungsmethode für autonome mobile Inspektionsroboter mit optischer Gasfernmesstechnik in industrieller Umgebung, Kassel 2015
Gero Bonow

Band 6: Dynamische Analyse großer verkoppelter Systeme mit Methoden der Komplexen Netzwerke am Beispiel des Inverse-Response-Verhaltens, Kassel 2015
Andreas Geiger

Band 7: Close range 3D thermography: real-time reconstruction of high fidelity 3D thermograms, Kassel 2018
Antonio Rafael Ordóñez Müller

Band 8: Zur regelungsorientierten Identifikation nichtlinearer Systeme mittels lokal affiner Takagi-Sugeno-Fuzzy-Modelle, Kassel 2019
Alexander Schrodt

Some novel approaches to estimate Nonlinear Output Error (NOE) models using TS fuzzy models for a class of nonlinear dynamic systems having variability in their outputs is presented in this dissertation. Instead of using unrealistic assumptions about uncertainty, the most common of which is normality, the proposed methodology tends to capture effects caused by the real uncertainty observed in the data. The methodology requires that the identification method must be repeated offline a number of times under similar conditions. This leads to multiple input-output time series from the underlying system. These time series are preprocessed using the techniques of statistics and probability theory to generate the envelopes of response at each time instant. By incorporating interval data in fuzzy modelling and using the theory of symbolic interval-valued data, a TS fuzzy model with interval antecedent and consequent parameters is obtained. The proposed identification algorithm provides for a model for predicting the center-valued response as well as envelopes as the measure of uncertainty in system output.

ISBN 978-3-7376-0650-9