ELECTRICITY PRICE TIME SERIES FORECASTING IN DEREGULATED MARKETS USING RECURRENT NEURAL NETWORK BASED APPROACHES

VISHAL SHARMA

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2011

Acknowledgements

It is a pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my Ph.D. supervisor, Assoc. Prof. Dipti Srinivasan. With her enthusiasm, her inspiration, and her great efforts helped to make neural networks and nonlinear theory fun for me. Throughout my thesis-writing period, she provided encouragement, sound advice, good teaching, good company, and lots of good ideas. I would have been lost without her support.

My warmest thanks and regards to the Power Systems Laboratory Officer Mr. Seow Heng Cheng for his helpful nature and dedication in making the laboratory such a nice place to work. I would also like to thanks Electrical Machines Laboratory Officer Mr. Woo Ying Chi. Without their support, it would have been impossible to carry out the research in the laboratory.

I am indebted to my many student colleagues, roommates and friends for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Vicky Lu SiYan, Atul Karande, Anupam Trivedi, Balaji Parasumanna Gokulan, Dr. Raju Kumar Gupta, Dr. Naran Pindoriya, Dr. Deepak Sharma, Dr. Yogesh Kumar Sharma, Sujit Kumar Barik and Ravi Tiwari. I would like to thank them for helping me get through the difficult times, and for all the emotional support, comraderie and caring they provided.

I wish to thank my brother, my sister in law and my beloved niece for providing a loving environment for me.

Lastly, and most importantly, I wish to thank my parents. I would have never reached so far in life without their constant love, support and encouragement. They bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

CONTENTS

ACKNOWLEDGEMENTS ……1

CONTENTS …….3

Summary ….....7

CHAPTER1

INTRODUCTION …… 18

CHAPTER 2

NEURAL NETWORKS …… 28

2.1 Learning in Neural Network …… 30

2.2 Stability of Neural Learning Algorithms …….. 33

2.3 Issues in NN Learning and Applications …….. 34

2.4 Implementation Example ……..40

2.4.1 Function Approximation …… 40

2.4.2 Pattern Classification …….. 42

2.5 Summary …….. 46

CHAPTER 3

DEREGULATED ELECTRICITY MARKETS AND VOLATILITY …… 47

3.1 Alternate Deregulation Models …… 49

3.2 Factors Affecting Volatility …….. 50

3.3 Models of Spot Prices …… 51

3.4 Market Design, Market Power and Pricing …….. 54

3.5 Summary …….. 59

CHAPTER 4

DYNAMIC CHARACTERISTICS OF ELECTRICITY PRICE TIME SERIES 60

4.1 Embedding Dimension 61

4.2 Fixed Point Characteristics 64

4.2.1 Locating Fixed Point 65

4.2.2 Dynamics in Neighborhood of Fixed Point 66

4.3 Lyapunov Exponents 67

4.4 Finite Time Lyapunov Exponent Analysis and Local Instability 68

4.5 Scale Dependent Lyapunov Exponent 71

4.6 Summary 74

CHAPTER 5

ELECTRICITY PRICE TIME SERIES PREDICTION USING RNN TRAINED USING INVARIANT DYNAMICS 76

5.1 Introduction 76

5.2 Weight Initialization 79

5.2.1 Identifying Fixed Point Location and Neighbourhood Dynamics 79

5.2.2 Fixed Point Based Initialization 82

5.3 Fixed Point Constraint During Learning 86

5.3.1 Extension to Nonlinear Constraint 89

5.4 Local Jacobian Learning 91

5.5 Summary 92

CHAPTER 6

ELECTRICITY PRICE TIME SERIES PREDICTION USING HYBRID RNN-FHN MODEL 94

6.1 Multiple Scale Dynamics in Electricity Price Time Series 95

6.2 Fitz-Hugh Nagumo Model 100

6.3 Proposed Model …… 101

6.4 Training of RNN in Hybrid Model …… 103

6.5 Prediction of Hourly Prices ……105

6.6 Training and Testing Data …… 106

6.7 Experimental Results …… 108

6.8 Interval Forecasting …… 112

6.9 Summary …… 118

CHAPTER 7

Multiscale Modelling of Electricity Price Time Series using Multi-Scale Neural Network .. 119

7.1 Slow-Fast Systems ……120

7.2 Multi-Scale Recurrent Neural Network (MSRNN) …… 123

7.3 MSRNN for Electricity Price Modeling …… 124

7.4 MSRNN Learning …… 127

7.5 Summary …… 130

CHAPTER 8

RESULTS AND DISCUSSION …… 131

8.1 Data and Preliminary Statistical Analysis …… 131

8.1.1 Data …… 131

8.1.2 Summary Statistics …… 133

8.2 Forecasting Indices Used …… 136

8.3 PGRNN Implementation Results …… 137

8.3.1 Results for PJM market …… 140

8.3.2 Results for Ontario market …… 143

8.3.3 Results for Victoria market …… 145

8.3.4 Results for NSW market 147

8.4 RNNFHN Implementation Results 149

8.4.1 Results for Ontario market 150

8.4.2 Results for PJM market 152

8.4.3 Results for Victoria market 155

8.4.4 Results for NSW market 157

8.5 MSRNN Implementation Results 159

8.5.1 Results for Ontario market 161

8.5.2 Results for PJM market 163

8.5.3 Results for Victoria market 165

8.5.4 Results for NSW market 167

8.6 Comparison of Performance of Three Proposed Models 170

8.7 Error Histogram Analysis 171

8.8 Discussion 175

8.9 Limitations of Developed Models 178

8.9.1 Limitation of PGRNN 178

8.9.2 Limitation of RNNFHN 178

8.9.3 Limitation of MSRNN 178

CHAPTER 9

CONCLUSION AND FUTURE WORK 180

9.1 List of Achievements 183

9.2 Future Work 184

REFERENCES 186

Summary

Electricity Price Time Series Forecasting in Deregulated Markets Using Recurrent Neural Network Based Approaches

In the past decade, electricity price time series system originating from recently deregulated electricity markets has been the focus of study for many researchers and power system engineers. These are complex dynamical systems which have tipping points at which sudden shifts to a spiking dynamical regime occurs. Although there are several techniques available for short term forecasting of electricity prices, very little has been done for accurate prediction of spikes along with otherwise volatile region of time series. High volatility and intermittent spikes are hallmarks of chaos taking place in electricity price time series. Modeling these systems require a dynamic approach with accurate approximation capabilities, such as recurrent neural networks. Recently recurrent neural networks have gained immense interest due to their unconventional ability to solve complex problems. However training them in complex dynamic environments such as electricity price time series is a challenging task due to various issues, which mainly include problem of local optima. However this problem can be rectified through intelligent learning of RNN incorporating heuristic knowledge of the system. Recently electricity price time series has been extensively investigated using nonlinear systems theory. Utilization of the extracted system invariant information to assist in solving issue of local optima can open a new dimension in recurrent neural network (RNN) learning and modeling. This thesis focuses on extraction of invariant dynamics of electricity price time series and incorporates them for developing RNN based pure as well as hybrid models for modeling electricity price time series and accurate prediction of price in spiking and nonspiking regime.

In this thesis, three RNN based approaches have been developed. First a novel recurrent neural network learning algorithm based on fixed point dynamics of time series system has been developed. This approach has been shown to bring the trained RNN model closer to exact nonlinear system. In the second approach, it has been proposed to hybridize the Recurrent Neural Network and a multi-scale excitable dynamic model to closely resemble the dynamic properties and spiking characteristics of time series system for obtaining an accurate forecasting model. This approach exploits the universal dynamic nonlinear approximation properties of RNN and spiking characteristics of self coupled FitzHugh Nagumo model. Fitz-HughNagumo (FHN) has been shown to exhibit dynamics close to electricity price due to presence of multiple scale dynamics. RNN trained using Evolutionary Strategies (ES) has been used for obtaining the parameter values of a coupled equation system (FHN). In third approach, the dynamic mechanism behind spike adding in time series has been extensively studied. Slow-fast dynamics and the corresponding complex homoclinic/heteroclinic scenarios, which are the underlying mechanism behind irregular spiking in time series have been exploited for modelling of multi-scale neural networks which are trained using singular perturbation theory and gradient descent algorithm. The developed models have been tested on various markets worldwide for different seasons. After extensive comparison with benchmarks, it has been demonstrated that the results are improved considerably.

To give an overview, the main contributions of this thesis are-

- Extraction of invariant measures of electricity price time series and confirm the presence of multiple scale dynamics in time series.

- Development of novel learning algorithm for RNN training incorporating invariant measures of time series.
- Development of a multi-scale neural network models and their learning algorithm employing singular perturbation theorem and use them for forecasting of price in deregulated electricity markets. The proposed approach improved prediction accuracy in spiking region.

LIST OF TABLES

Table 2.1	Performance of RM Model With Gradient Descent as Order of Polynomial is Varied ……	44
Table 2.2	Comparison of Performance Measure of Various Algorithms ……..	45
Table 4.1	Embedding Dimension Calculation for Various Deregulated Markets ……..	63
Table 4.2	Eigenspectrum of Fixed Point of Time Series ……	66
Table 4.3	Lyapunov Exponent for Various Deregulated Markets ……	68
Table 6.1	Evolutionary Strategy Parameters ……	105
Table 6.2	Testing Dataset Used ……	106
Table 6.3	Performance Comparison over Deregulated Markets ……	111
Table 6.4	Interval Forecast ……	116
Table 6.5	Error Statistics ……	117
Table 8.1	Testing Dataset Used ……	133
Table 8.2	Statistics of Training Data ……	134
Table 8.3	Prediction Error for Worldwide Deregulated Markets Using Three Proposed Approaches of RNN Learning ……	138
Table 8.4	Comparison of Prediction Error for proposed PGRNN trained using combination of all three approaches with Benchmarks ……	139

Table 8.5	Comparison of Proposed RNNFHN and Benchmark Approaches for Worldwide Markets ……	150
Table 8.6	Comparison of Proposed MSRNN and Benchmark Methods on Worldwide Deregulated Markets …….	160
Table 8.7	Comparison of the Proposed Approaches for Worldwide Deregulated Markets …….	170
Table 8.8	Error Histogram Plots for Ontario Market …….	172
Table 8.9	Error Histogram Plots for PJM Market ……..	173
Table 8.10	Error Histogram Plots for NSW Market ……	174
Table 9.1	Comparison of the Proposed Approaches for Worldwide Deregulated Markets …….	182

LIST OF FIGURES

Figure 2.1	Schematic of a Biological Neuron …….	29
Figure 2.2	Multi Layer Perceptron …….	30
Figure 2.3	Predicted and Actual Data for the Load Forecasting Problem ……	41
Figure 2.4	Variation of DR and FAR as classification threshold is changed ……	46
Figure 4.1	2D Projection of phase space of Ontario time series …….	60
Figure 4.2	Recurrence Plot of Ontario time series ……..	61
Figure 4.3	Fractal Dimension plots for California market in the non spiking region of time series for embedding dimensions 4 to 22 are shown for even numbers. Fractal dimension is estimated as slope …….	63
Figure 4.4	Third Largest FTLE for Ontario Market Time Series ……	70
Figure 4.5	SDLE plot for Ontario Market ……	74
Figure 5.1	Elman State Feedback Recurrent Neural Network …….	78
Figure 6.1	Smooth Orthogonal Values of slow subsystem ……	98
Figure 6.2	Dominating Slow Component vs Time Series ……	99
Figure 6.3	Time response of FHN for ε=0.11, k=1, λ=1, b=2, γ=1, β=1, A=2, ωf=2 and I=1 ……	101
Figure 6.4	Proposed Hybrid Model with RNN and FHN coupled system …….	102
Figure 6.5	Chromosome Representation …….	104
Figure 6.6	Selection of range of normalization for spiky time series. The high value spikes are neglected ……..	107

Figure 6.7 Actual and Predicted curve in Spanish market during Summer …….. 109

Figure 6.7 Actual and Predicted curve in Ontario market during Spring …… 109

Figure 6.8 Enlarged section of Interval Forecasts for Ontario market ……. 113

Figure 6.8 Enlarged section of Interval Forecasts for California market with large forecasting range around spiky region ……. 113

Figure 6.9 Histogram for error values of California market ……. 117

Figure 7.1 Multiple Scale Neural Network …….. 126

Figure 8.1 Actual and Predicted curve in PJM market during Week1 for PGRNN ……. 141

Figure 8.2 Actual and Predicted curve in PJM market during Week2 for PGRNN ……. 141

Figure 8.3 Actual and Predicted curve in PJM market during Week3 for PGRNN ……. 142

Figure 8.4 Actual and Predicted curve in PJM market during Week4 for PGRNN ……. 142

Figure 8.5 Actual and Predicted curve in Ontario market during Week1 for PGRNN …. 143

Figure 8.6 Actual and Predicted curve in Ontario market during Week2 for PGRNN …. 144

Figure 8.7 Actual and Predicted curve in Ontario market during Week3 for PGRNN …. 144

Figure 8.8 Actual and Predicted curve in Victoria market during Week1 for PGRNN …. 145

Figure 8.9 Actual and Predicted curve in Victoria market during Week2 for PGRNN …. 146

Figure 8.10 Actual and Predicted curve in Victoria market during Week3 for PGRNN …. 146

Figure 8.11 Actual and Predicted curve in NSW market during Week1 for PGRNN …….. 147

Figure 8.12 Actual and Predicted curve in NSW market during Week2 for PGRNN …….. 148

Figure 8.13 Actual and Predicted curve in NSW market during Week3 for PGRNN …….. 148

Figure 8.14 Actual and Predicted curve in NSW market during Week4 for PGRNN …….. 149

Figure 8.15 Actual and Predicted curve in PJM market during Week1 for RNNFHN …….. 151

Figure 8.16 Actual and Predicted curve in PJM market during Week2 for RNNFHN …….. 151

Figure 8.17 Actual and Predicted curve in PJM market during Week3 for RNNFHN …….. 152

Figure 8.18 Actual and Predicted curve in PJM market during Week4 for RNNFHN …….. 153

Figure 8.19 Actual and Predicted curve in Ontario market during Week1 for RNNFHN .. 153

Figure 8.20 Actual and Predicted curve in Ontario market during Week2 for RNNFHN .. 154

Figure 8.21 Actual and Predicted curve in Ontario market during Week3 for RNNFHN .. 154

Figure 8.22 Actual and Predicted curve in Victoria market during Week1 for RNNFHN .. 155

Figure 8.23 Actual and Predicted curve in Victoria market during Week2 for RNNFHN .. 156

Figure 8.24 Actual and Predicted curve in Victoria market during Week3 for RNNFHN .. 156

Figure 8.25 Actual and Predicted curve in NSW market during Week1 for RNNFHN ……. 157

Figure 8.26 Actual and Predicted curve in NSW market during Week2 for RNNFHN ……. 158

Figure 8.27 Actual and Predicted curve in NSW market during Week3 for RNNFHN ……. 158

Figure 8.28 Actual and Predicted curve in NSW market during Week4 for RNNFHN ……. 159

Figure 8.29 Actual and Predicted curve in PJM market during Week1 for MSRNN ……. 161

Figure 8.30 Actual and Predicted curve in PJM market during Week2 for MSRNN ……. 162

Figure 8.31 Actual and Predicted curve in PJM market during Week3 for MSRNN ……. 162

Figure 8.32 Actual and Predicted curve in PJM market during Week4 for MSRNN ……. 163

Figure 8.33 Actual and Predicted curve in Ontario market during Week1 for MSRNN … 164

Figure 8.34 Actual and Predicted curve in Ontario market during Week2 for MSRNN … 164

Figure 8.35 Actual and Predicted curve in Ontario market during Week3 for MSRNN … 165

Figure 8.36 Actual and Predicted curve in Victoria market during Week1 for MSRNN … 166

Figure 8.37 Actual and Predicted curve in Victoria market during Week2 for MSRNN … 166

Figure 8.38 Actual and Predicted curve in Victoria market during Week3 for MSRNN … 167

Figure 8.39 Actual and Predicted curve in NSW market during Week1 for MSRNN …….. 168

Figure 8.40 Actual and Predicted curve in NSW market during Week2 for MSRNN …….. 168

Figure 8.41 Actual and Predicted curve in NSW market during Week3 for MSRNN …….. 169

Figure 8.42 Actual and Predicted curve in NSW market during Week4 for MSRNN …….. 169

LIST OF PUBLICATIONS

International Journals

- D. Srinivasan, Vishal Sharma, K.A. Toh, "Reduced multivariate polynomial based neural network for automated traffic incident," *Neural Networks*, vol.21, pp. 484-492, March-April 2008.

- D. Srinivasan, Saptak Sanyal and Vishal Sharma, "Freeway incident detection using hybrid fuzzy neural network," in *IET Intelligent Transport Systems*, vol.1, pp.249-259, 2008.

- Vishal Sharma and D. Srinivasan, "Novel method of recurrent neural networks learning using invariant features of time series," in *IEEE Transactions on Neural Network*, accepted.

- Vishal Sharma and D. Srinivasan, "Price time series forecasting in deregulated power markets using multi-scale neural networks," in *IEEE Transactions on Power Systems*, under review.
- Vishal Sharma and D. Srinivasan, "Day ahead price forecasting in deregulated power markets using hybrid RNN-FHN model," in *Engineering Applications of Artificial Intelligence*, Accepted.

- Vishal Sharma and D. Srinivasan, "Dynamic analysis of electricity price time series in deregulated markets," in *Electric Power Systems Research*, under review.

International Conferences

- Vishal Sharma and D. Srinivasan, "Hybrid Model Incorporating Multiple Scale Dynamics for Time Series Forecasting," *IEEE International Joint Conference on Neural Networks*, 2011, under publication.

- Vishal Sharma and D. Srinivasan, "Spiking Neural Network Based on Temporal Encoding for Electricity Price Time Series Forecasting in Deregulated Markets," *IEEE International Joint Conference on Neural Networks*, pp 1-8, 2010.

- Vishal Sharma and D. Srinivasan, "Evolutionary Computation and Economic Time Series Forecasting," *IEEE Congress on Computational Intelligence*, pp 188-195, 2007.

- Dipti Srinivasan and Vishal Sharma, "A Reduced Multivariate Polynomial Based Neural Network Model Pattern Classifier for Freeway Incident," *IEEE International Joint Conference on Neural Networks*, pp 1-8, 2007.

Chapter 1

Introduction

This thesis focuses on developing a better understanding of spike mechanism in electricity price time series in deregulated markets and develop novel recurrent neural network (RNN) based models and their learning algorithms to improve the prediction on deterministic time series system. This approach can also be seen as attaining heuristic information about the system in order to achieve global optimal solution in recurrent neural networks learning for modeling the complex time series system. The objectives of thesis can be stated as-

- To study deregulated markets, price formation mechanism and factors affecting volatility of price.
- Analyze electricity price time series from nonlinear theory perspective and understand the underlying dynamics of chaotic and spiking behavior in time series.
- To employ the obtained information as heuristics to develop recurrent neural network based models and their learning algorithms for accurate prediction of electricity prices.

The recent deregulation of electricity markets is one of the major issues in power system studies. This current trend, which is increasing further in worldwide markets, has led to competition and created opportunities for various market participants to trade electricity. Electricity price is determined using a bidding based system, where the final price is the outcome of a complex process which depends on various intrinsic and extrinsic factors. Moreover non-storable nature of electricity as a commodity causes state of disequilibrium in demand and supply. Trading in electricity markets is a more challenging task compared to traditional financial trading, as the former exhibits higher volatility. Electricity price differs

from other financial commodities as it lacks storage and all generated electricity must be consumed, which creates new scope of volatility. In this scenario, it is interesting to examine the dynamic nature of resultant electricity price time series and the scope of predictability in this system. Moreover, price forecasting is an important requirement for deriving power bidding strategies for profit maximization. Accurate estimation of future base load price is extremely helpful for the producers and consumers for deriving their respective bidding strategy and risk management. The price time series exhibits intermittent spikes at varying locations with varying intensity. It is essential to have a good approximation throughout the series, however particular attention should be paid to these risky spiky regions, failing which can result in loss of millions of worth of utilities.

1.1 Literature Review on Electricity Price Modeling

Until now several approaches have been proposed for time series modeling. These applied methodologies can be broadly divided into five broad classes. First there are *production-cost balance* based approaches which simulate the generation operations and aim to balance demand at minimum cost. However these approaches ignore strategies in bidding based scenarios hence not applicable for deregulated competitive markets. There are *equilibrium* based approaches which consider strategic bidding based scenarios. They provide excellent insight into prices above marginal cost and their influence on decision of market players. However, there is risk involved as the strategy and risk behavior of market players has to be decided upfront. The cournot nash equilibrium [1] framework of equilibrium approaches which provide price higher than real prices, has limited applicability in everyday market. *Fundamental* methods describe the price dynamics using impact of importance of economic and physical factors on electricity price. In these methods the association between load, weather situations and other fundamental drivers are postulated

and the input units are predicted using parametric or nonparametric approaches [2]. However due to fundamental nature of data collected over long time ranges, these models are suited for medium and long term rather than short term range predictions. On the other hand there is class of *quantitative* models which characterize and employ statistical properties of electricity price while the ultimate objective is evaluation of derivatives and management of risk. These models obtain the main characteristics of electricity price time series at short horizons instead of actual modeling and prediction. Although these models are simple and analytically tractable, their former feature is a limitation. Another class of models constitutes *statistical* approaches which aim to obtain optimal model in term of forecasting indices. Either these models employ econometric models in context of power market or statistical models developed for load forecasting. Most widely used statistical models include multivariate regression, smoothing models and time series based techniques. These models can incorporate fundamental factors such as load demand and fuel prices to enhance their performance. Statistical models include parametric approaches like ARIMA and their higher order variants [3-6], or hybrid models combining statistical modeling with the basic concept of supply-demand balance [7]. Electricity demand is heavily influenced by economic and business activities and by the weather. Demand is usually characterized as highly inelastic because it is a necessary commodity [8]. When there are low levels of demand, generators supply electricity using base-load units with low marginal costs, but, during summer and winter seasons, during certain days of the week and even within the day, higher quantities are needed and generators with higher marginal costs enter into the system. Such seasonal factors have been studied in Ref. [9-10]

Since increases in demand push up prices, there are increasing incentives of even expensive generators to enter the supply side, so that some degree of mean-reversion is

expected in prices. Most studies employ mean-reverting models [11-14], although some allow for non-mean-reverting behavior [15]. Some show that there are interesting interactions on the degree of mean-reversion in the price process with other features such as time-varying conditional volatility and price spikes [14]. Ref.[14], specifically, incorporates multiple jumps, regime-switching and stochastic volatility into a number of mean-reversion models and show how sensitive real-option-based models of physical assets in generation and transmission can be.

Volatility in electricity prices varies over time with weather-related and other demand and supply forces and it is likely mean-reverting itself for similar reasons as outlined above. Ref. [16] proposes a mean-reverting proportional volatility model and find empirical support with intraday prices over constant volatility, geometric Brownian motion models. Ref. [17] proposed ARCH models for heating oil, natural gas, crude oil and electricity prices but show the limitations of the functional model for electricity, which show results close to integrated (non-mean-reverting, or "explosive") processes for volatility. Authors in ref. [9, 13] use generalized ARCH (GARCH) models for electricity prices and ensure stationarity in volatility when price spikes are captured by separate jump-diffusion processes. Ref. [14] also emphasizes the importance of modeling jump processes in electricity prices, especially as they relate to monthly seasonal factors.

These models perform quite well during nonspiking regions where the seasonality trends prevail. Finally there are nonparametric *artificial intelligence* based techniques such as ANN, data mining expert models, SVM, fuzzy logic etc [18-20]. AI based models tend to be flexible and can handle complexity and non-linearity. This makes them promising for short-term predictions. In fact a number of authors have reported their excellent performance in price forecasting [18-19]. These models have the advantage of tractability which is

important for modeling electricity markets but lack performance accuracy in few cases. Most of the developed models compare their performances with ANN and statistical approaches. However, some of the works indicate that there might be serious problems with the efficiency of ANNs and AI-based methods in general. The key limitation comes from the difficulty in successful training of models.

Very few of proposed approaches have explicitly addressed the problem of modeling spikes. Some authors have used data mining approach to model spikes uncertainty, their level and the associated confidence interval, while wavelet-neural cascade technique for normal level prices. Bayesian expert model with support vector machines (SVM) for spike forecasting has also been adopted. One of the most researched strand of models for spike modeling are stochastic regime switching models which carefully identify different dynamics as different operating regimes of the dynamic model and employ probabilistic transition among them. A very complete and recent example in this attempt is given in ref. [21] where authors have applied the regime switching model as the describing dynamic model. These models incorporate various system characteristics and governing factors which lead to more robust and realistic spot price modeling, as well as bringing insights about the spot price dynamics under various supply and demand scenarios. While not a forecasting tool, the dynamics of these models behave as expected and match the statistical properties of the time series with accuracy.

1.2 Theoretical Overview and Motivation

Spike formation is a well studied phenomenon in many works where it has been proven to be a deterministic event. It has been proven that the location of spikes in dynamic systems can be determined from the dynamics on the invariant manifold [22]. Moreover,

the intensity and amplitude of spike is related to the hyperbolicity of the manifold in the neighborhood of spike location and thus is a deterministic quantity if a well trained nonlinear model is employed [22]. Neural networks have been proven to be universal approximators of nonlinear dynamic systems [23]. In this work neural networks based models and their variants have been employed for electricity price time series modeling. A brief review of neural networks and associated modeling issues is given in chapter 2 where steps by step approach towards acquaintance with neural networks and their implementation.

The spiking behaviour in electricity prices is intermittent at varying locations with varying intensity. From dynamical systems point of view, it represents a very complex system with possibly multiple behaviours in same system with possible varying stability along trajectory in phase space. While processing and forecasting time series of such system, there are many issues one of which involves ability to capture the exact nonlinear features of dynamic system represented by time series. This becomes absolutely necessary when the system exhibits critical transitions. Most of the real world physical system have critical thresholds, also known as tipping points, at which the system abruptly shifts its state from one to another. This phenomenon is readily seen in medicine, weather, finance etc. In earth, abrupt shifts in ocean currents may cause climate changes [24-25]. Spontaneous changes known as epileptic seizures can occur in brain EEG signals [26]. Although it is of utmost importance to predict these changes, the prediction becomes notably hard because the system shows unnoticeable change before tipping point is reached. Intermittent spiking is one such phenomenon. The objective is to understand the basic mechanism underlying the states exhibited by the system and using the same to model it. The study of deregulated markets and the underlying price formation and spiking mechanism is a crucial issue which

deserves a separate study. In chapter 3 we analyze the structure of deregulated electricity markets worldwide, price formation mechanism and the factors responsible behind volatile nature of electricity price time series. Firstly the mechanism behind electricity price in deregulated market is discussed. The deformation of structure of market, due to deregulation, is studied and various factors behind complexity of electricity price behaviour are identified. The studies reveal that the structure of various deregulated markets at different geographical locations is different and there are different sources responsible for complexity in time series. However they share one similar spiking attribute. They exhibit spike occurring at irregular time intervals and exhibit chaotic oscillations in the nonspiking region. This motivates the study of the chaotic features of the phase space embodying the time series.

Different Chaotic systems have strange attractors characterized by properties of the attractor independent of particular trajectory which are called invariants of the system. The study of dynamic characteristics of this kind of time series include study of invariant sets of time series and, for this particular work, extraction of dynamic attributes which are the key to understanding and modelling of neural networks based on time series. The invariant set of a dynamical system is a general entity in nonlinear dynamics. It is imperative to analyze time series using nonlinear systems theory and observe the invariants measures constituting time series. Embedding dimension is the basic invariant measure which is crucial for reconstruction of phase space of time series. Also, it is imperative to measure Lyapunov exponent of time series to measure the degree of chaos in the system. However due to presence of possible intermittent variations (spiking) in dynamics, Lyapunov exponent is not a reliable index, which is why Finite Size Lyapunov Exponent (FSLE) and Scale Dependent Lyapunov Exponent (SDLE) are observed to analyze the transient dynamics of time series

along with global behaviour. Observing FSLE allows detecting the possible presence of hyperchaos in the system and loss of local hyperbolicity. SDLE analysis allows establishing the fact that electricity price is not a noisy time series. On the other hand it exhibits complex irregular behavior due to the presence of dynamics on multiple scales. Fixed point is also one of the most important invariant of time series. Most of the real world time series consist of saddle fixed points in phase space giving rise to complicated homoclinic or heteroclinic scenarios. Eigenvalues of these saddles determine the behaviour of the system. From dynamical system perspective, spiking transition can be approached from variation in stability in phase space, synchronization in coupled systems, multiple scale behaviour, etc. For the purpose of approximating this complex system, a dynamic variant of neural network, i.e. Recurrent Neural Network (RNN) is used in this work. RNN are universal dynamic system approximators, which allow a wide variety of dynamic behaviour [27]. As mentioned earlier, in this work a dynamical system approach based on nonlinear neural network is adopted to model spiking and normal dynamics of time series. The calculated invariant features of time series have been exploited for their modelling. The fixed point dynamics and FSLE are used for RNN weight initialization and learning. In order to achieve closer approximation of nonlinear dynamics of time series, we trained a pure state feedback recurrent neural network using the calculated invariant measures. It is shown that incorporation of invariant measures in the learning process results in better learning of time series. However the key observation in this work is the presence of dynamics on multiple scales. In the later part of this work, the multiple scale dynamics approach is adopted in Chapter 6. The spiking behaviour is described as critical transition in a multiple scale system where the system dynamics bifurcate due to variation in “parameter”. One of the dynamical systems exhibiting multiple scale dynamics is used to study spiking, the well known Fitz-Hugh

Nagumo (FHN). A dynamic system with slow and fast scales, namely Fitz-Hugh Nagumo (FHN), is used and hybridized with recurrent neural networks. The property of the multiple scale equation system allows the mechanism of spiking in such regimes. In turn, the parameters and coupling variables of this excitable system are determined using an RNN based model. As a result the learned hybrid model would achieve a desired level of modelling accuracy. The developed hybrid model was tested in various markets worldwide over different seasons to test its forecasting ability, adaptability and robustness. Most volatile electricity markets, California, Australia, PJM, Spain and Ontario market in Canada were modelled using the proposed approach. Extensive comparative studies suggest that our approach yields favourite results in hour-ahead and day-ahead market.

For more accurate modeling of time series and the associated slow fast dynamics, a multiple scale neural network (MSNN) is developed in chapter 7. Slow-fast systems deal with slow manifold and fast manifold where the key dynamics of time series occur on fast invariant manifold while the dynamics occurring on slow manifold is responsible for intermittent critical transitions. The developed model is trained using singular perturbation theory for slow-fast systems combined with gradient descent algorithm. The homoclinic scenarios involved behind spike adding mechanism are identified and employed for modeling of proposed model.

1.3 Structure of Thesis

This thesis is organized as follows. In the next chapter, a brief overview of neural networks and the associated modelling issues is presented. In chapter 3, the effect of deregulation on electricity markets and the mechanism behind price generation are discussed. Various factors affecting behaviour of price have been discussed and basic

statistical properties have been studied. In chapter 4 the nonlinear dynamic characteristics of time series have been extracted and analyzed. The invariant attributes of the underlying nonlinear system are extracted which are later exploited for dynamic information based modelling of RNN. These include Lyapunov exponent, finite time Lyapunov exponent, multiple scale behaviour, embedding dimension and fixed point dynamics. Moreover the deterministic multiscale nature of time series is established and it is proved that electricity price is not a stochastic variable. In chapter 5, the dynamic attributes of time series extracted in chapter 3 are incorporated in modelling recurrent neural networks. In chapter 6 and 7, the multiple scale dynamics of time series have been exploited. Chapter 6 briefly describes behaviour of FHN in slow and fast time scales and uses RNN to modulate FHN for accurate prediction in time series. In chapter 7, multiscale (slow fast) dynamics of time series are extensively studied and the applications of multiple scale recurrent neural networks are proposed. The implementation results of these developed models have been given in chapter 8 along with discussions. In chapter 9, conclusion and future work have been given followed by references.

Chapter 2

Neural Networks

In this chapter a brief introduction of neural network has been given. Various issues have been discussed which require major attention while modeling neural networks and developing their learning algorithms. Later in the chapter, introductory implementation examples are given in order to provide a simple neural network modeling overview.

In past many years, the advancement of powerful computing systems allowed advancement in research in field of neural networks. A neural network is a representation of model of biological networks in brain and is a conceptual circuit capable of performing computational task. Brain analyzes all patterns of signals sent, and from that it interprets the type of information received. The basic model is founded based on biological neural network in brain. In neuroscience, a neural network describes a population of physically interconnected neurons or a group of disparate neurons whose inputs or signaling targets define a recognizable circuit. Communication between neurons often involves an electrochemical process. The interface through which they interact with surrounding neurons usually consists of several dendrites, which are connected via synapses to other neurons, and one axon (output connection). If the sum of the input signals surpasses a certain threshold, the neuron sends an action potential [23].

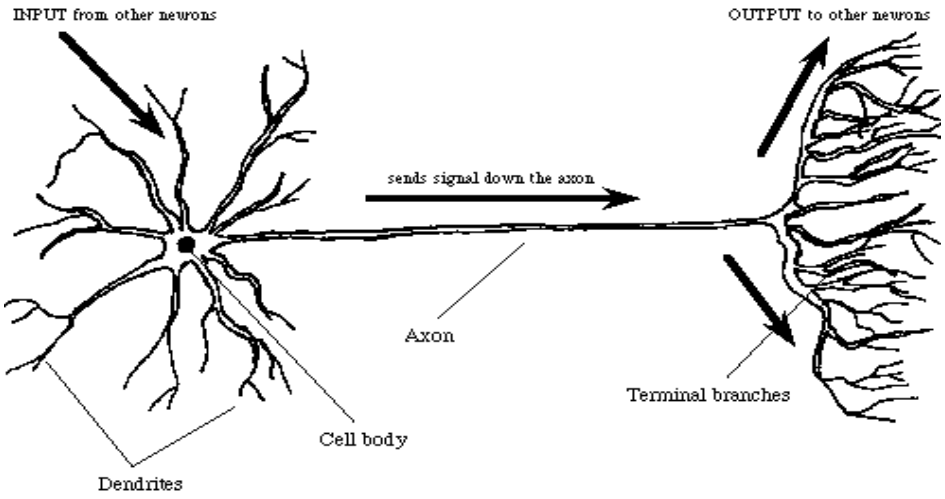


Fig 2.1 Schematic of a Biological Neuron

An artificial neural network (ANN), or commonly just neural network (NN) is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing. An artificial neural network involves a network of simple processing elements (artificial neurons) which can exhibit complex global behavior, determined by the connections between the processing elements and element parameters. Neural Networks are nonlinear structures. The utility of artificial neural network models lies in the fact that they can be used to infer an input output functional relationship from observations and also to use it. This is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical. An artificial neural network is a system based on operation of biological neural networks and thus is an emulation of biological neural networks. It is an adaptive system, by which it means that the parameters change during operation [23].

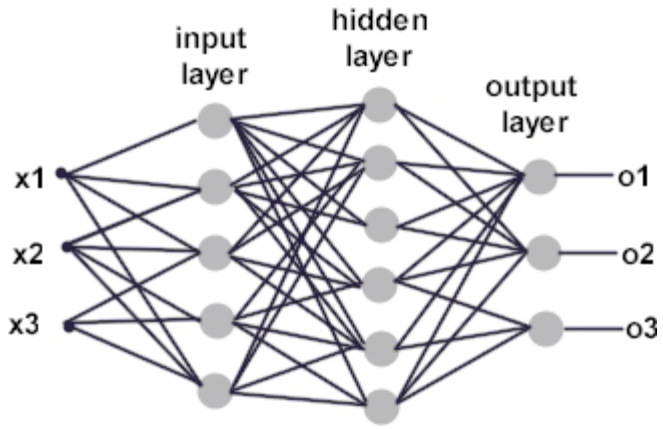


Fig 2.2 Multi Layer Perceptron

An artificial neural network is developed to perform two main functions, pattern recognition and function approximation. The problem of electricity price forecasting falls under domain of function approximation. The problem of classification involves task of pattern recognition to assign an input pattern to one of many possible classes. This involves extensive application of algorithmic implementations such as associative memory. The task of function approximation is to approximate an unknown function subject to noise, given its attributes. Various streams of engineering require function approximation. In most cases, such as time series of dynamical system, the approximated function is required to be capable of task of prediction of future values in the time sequence data. In other words time series prediction is function approximation with the consideration of time factor. The system is dynamic which implies that the same set of inputs can produce different results depending on the current state. In the next few sub sections various issues related to neural network implementation are discussed.

2.1 Learning in Neural Networks

Learning is a process by which the weight parameters and other free parameters of a neural network are adapted through a process of simulation by the environment in which the network is embedded [23]. Learning and generalization is perhaps the most important topic in neural network research [28], [29], [30]. Learning is the ability to approximate the underlying behavior adaptively from the training data while generalization is the ability to predict well beyond the training data. A number of practical network design issues related to learning and generalization include network size, sample size, model selection, and feature selection have been studied extensively in the past [31-44]. The type of learning is determined by the manner in which the parameters are changed. A prescribed set of well defined rules for the solution of a learning problem is called a learning algorithm. The types of learning algorithms can be categorized according to the manner in which the parameter updating takes place. These can be broadly stated as-

1. Error Correction Learning – Here the training algorithms make use of difference between the desired output and the output signal produced by the neural network. The error term actuates the update mechanism, the purpose of which is to apply a sequence of corrective adjustments to the synaptic weight of the neuron of the network.

2. Memory-Based Learning – In memory based learning, all of the past experiences are explicitly stored in a large memory of correctly classified input-output examples: {(xi, di)}, where di denotes the desired response corresponding to an input vector xi. The key ingredients of memory based learning rules are-

a) Criterion used for defining the local neighborhood of the test pattern vector.

b) Learning rule applied to the training examples in the local neighborhood of the test pattern vector.

One of the classic examples of memory-based learning rule is the Nearest Neighbour learning rule.

3. Hebbian Learning rule – Hebbian learning rule is one of the most famous and oldest of all learning rules. From the Neurobiological point of view, we can briefly explain this learning rule as- If two neurons on either side of connection are activated simultaneously (i.e. synchronously), then the strength of that synapse is selectively increased. While if two neurons on either side of a connection are activated asynchronously, then that connection is selectively weakened or eliminated.

4. Competitive Learning – In competitive learning, the output neurons of a neural network compete among themselves to get active (fired). While in hebbian learning, several output neurons may be activated at the same time, here only single output neuron is active at one time. This feature makes them suitable to discover statistically salient features that may be used to classify a set of input patterns.

One other important way of classifying the learning procedures is to classify them as teacher Forced learning algorithms and algorithms without teacher forced learning. In teacher forced learning or Supervised Learning, teacher may be thought of as having knowledge of the environment, with that knowledge being represented by a set of input-output examples. The network parameters are adjusted under the combined influence of the training vector and error signal. This adjustment is carried out in a step-by-step fashion with the aim of eventually making the neural network to emulate the teacher. However in the paradigm known as learning without a teacher, as the name implies, there is no teacher to oversee the learning process. Two subdivisions can be identified under this paradigm –

a) Reinforcement Learning – In Reinforcement learning, which is closely related

to dynamic programming, the learning of an input-output mapping is performed through continued interaction with the environment in order to minimize a scalar performance index.

b) Unsupervised Learning – In unsupervised or self-organized learning there is no external teacher or critic to oversee the learning process. Provision is made for a task-independent measure of the quality of representation that the network is required to learn, and the free parameters are optimized with respect to that measure.

2.2 Stability of Neural Learning Algorithms

Engineering applications of neural networks rely crucially on qualitative properties of stability and dynamic behaviors of the networks. The existence and convergence of a unique equilibrium point are of importance for a neural network. The network should have a unique global attractive equilibrium point, where uniqueness of the equilibrium point is required to avoid the risk of spurious response or the common problem of local minima and hence ensure global optimization. In the past few years, stability property of neural networks with delays or without delays has been also extensively studied by many researchers, and a large number of stability criteria have been derived. Some of criteria have been derived for the global asymptotic stability in [45-52], for global exponential stability in [53-57] and for absolute stability in [58]. Among the above three kinds of stability properties, global exponential stability have the best character. For example, in designing a neural networks, one concerns not only on the stability of the system but also on the convergence rate, that

is to say, one usually desires a fast response in the neural network, so it is important to determine the exponential stability. Ref [59] discusses the network with a unique equilibrium point which is globally exponentially stable.

2.3 Issues in NN learning and applications

In this work, we have mainly focused on supervised error correction learning. In this category, the most famous and widely used algorithm is the gradient based back propagation (BP) rule of learning. The advantage of back propagation based learning lies in the simplicity in learning and implementation. But the algorithms developed have to deal with various issues, for the improvement of which, various modifications have been introduced. Some of the issues are discussed and some recent developments towards handling of those issues are given below –

a) Over Fitting: Sufficient amount of data is required for the effective training of Neural Network structures. Also the architecture of the network should be chosen intelligently so that the modeling of nonlinear function is appropriate. If the network has less number of approximating units, the network is not able to capture the features of the function and thus function is under fitted. If the network size is chosen larger, the network overfits the training data approximating the noise also. In recent years, many ANN-based forecasters are proposed for learning the highly nonlinear load pattern, yet their effectiveness are limited by the reduction of training data, which causes these ANN models to be susceptible to "over-fitting". "**Over-fitting**" is a common ANN problem that describes the situation that the model memorizes the training data but fails to generalize well to new data.

Ref. [60] discusses the problem of “over-fitting” and some common generalization learning techniques in the ANN literature, as well as introducing a new Genetic Algorithm-based regularization method.

b) Local Optima: Neural network learning is a multi-modal nonlinear optimization with many local minima. One of the main limitations of BP based algorithm in most cases is the inability of algorithm to escape local optimal solution in case of complex environment where the error space is multimodal [61]. Researchers have worked over this problem over years. To overcome the deficiencies of local-search methods, global minimization methods have been developed which can be classified into probabilistic and deterministic. These include *covering methods* which detect the regions not containing global minima and exclude them from consideration during learning [62], *generalized gradient descent* methods which flatten the search regions containing local optima [63], *clustering methods* which prevent redetermination of already known local optima, *random search methods* such as evolutionary algorithms [64-65] and stochastic models which use random variables to model unknown values of objective function [66]. The deterministic methods do not work well when the search space is too large for the deterministic methods to cover the search space adequately. On the other hand, propbabilistic global minimization methods use probability for decision making. The simplest of them use restart the search process to bring out the search from local optima. All probabilistic methods are weak in either local or global search. All these mentioned approaches at best find good local minima of multimodal function. In

exceptions with restrictive assumptions such as Lipschitz condition, accuracy of algorithm performance is guaranteed. Usually in the best performing scenarios, global optimization is achieved using hybrid models which use heuristic optimization method [67] (which is technically one of the ideas behind approach adopted in this thesis). In [67], low-discrepancy sequence of points and a simplex local search is combined for achieving global optimal solution. One of the recent examples is [68] in which the algorithm used for training of weight vectors in a simple single hidden layer Neural Network is different from BP. This algorithm resolves this issue by following steps based on Lyapunov stability theory. The error function defined in this algorithm is UNIMODAL. The basic steps followed are: Define error energy lyapunov function V(k) (Positive Definite) of tracking error between output of neural network and desired responses. Then choose the adaptive weight updating rule such that $V(k)-V(k-1) < 0$. This ensures minimization of error energy. In one of the recent examples [69], Neural Network training is done using Chaotic PSO (Particle Swarm Optimization) based algorithm. The reason behind adding chaotic perturbations to PSO is to improve the exploration capabilities. This trained network is used for the control of nonlinear systems which are difficult to model mathematically because of strong nonlinearities or lack of information. The unique feature of chaotic process is that it can traverse every state by its own dynamics and each state is traversed only once. So the chances of trapping in local optima are minimal. The chaotic system used is a tent-map system.

c) Non Stationary Environment: This is one of the key issues which are faced in the case of complex problems where the nature of the environment is time varying. Thus the underlying dynamics of the time series are nonstationary and varying. Some of the texts have tried to deal with this problem. For ex. In [70], the problem of temporal pattern recognition is handled in a time varying environment. The problem in handling with temporal pattern recognition is the appropriate representation of time-varying patterns. As temporal patterns evolve, the earlier trained data based on older patterns is not able to classify them appropriately.

d) Unavailability of Information on Network Structure: Neural Network training and its application is associated with several other limitations which had been investigated in the past. One of the disadvantages is that the exact architecture, which includes number of layers, number of neurons in each layer and activation function to be used, of the network which will give the best results is unknown in most of the applications. In most of the cases, the architecture of the network is randomly tested and the best architecture is chosen after several experimentations. This process is time consuming. Some of the literatures have tried to solve this problem. For example in a recent effort [71], simultaneous training of network architecture and weights is done with tabu search and Simulated Annealing used separately as training algorithms. For this training procedure, cost is a function of both training error and network size. In another recent example [72], the number of neurons in the hidden layer of a 3 layer neural network is calculated for a classification problem. This paper tells the importance of number of units in

the hidden layer of an MLP. It further uses Singular Value Decomposition (SVD) for determining the optimal number of neurons in the hidden layer. The basic idea uses classification as basis. The optimal number of hyperplanes needed for complete classification in higher dimension space is related to linear independency of patterns in that space. Thus the number of hidden neurons will be equal to the rank of the correlation matrix. But the actual data is noisy. So using this method will produce more number of neurons than required which will cause overfitting. This problem is resolved by observing the eigenvalues obtained by SVD. Various threshold level based criteria (applied on the smallest eigenvalues) are proposed in this paper to remove noise factor. One other limitation is seen while training of Neural Networks for problems with high dimensions.

e) High Dimensional Data: In the problem environment where the dimension of the data used is very high, the input layer size is larger which results in requirement of larger number of neurons in subsequent layers of the networks. The training becomes a complicated task. One of the solutions to this problem is feature selection whereby a data space in high dimension is transformed into a feature space. The transformation is designed in such a way that the data set may be represented by a reduced number of effective features and yet retaining the most of the intrinsic information or the data undergoes dimensionality reduction. Principal Component Analysis (PCA) is one of the techniques which maximize this dimensionality reduction in the most effective manner. A lot of research has been done on this technique and its variations and improvements. One of the examples is [73] where

Minor Component Analysis is introduced and used. Minor Component Analysis (MCA) is the converse of PCA. It uses the eigenvector associated with the smallest eigenvalue of the input covariance matrix. The information of covariance matrix is not needed in advance. This algorithm sees the associated Neural Network as a stochastic Discrete-time system for which the stability theories described by ODEs do not work. This reference proves the stability of this deterministic discrete time system using conditional expectation of weight vector. The goal of MCA is to converge the weight vector to the smallest eigenvector of the covariance matrix using MCA learning law. The convergence analysis is proved in the text. One of the other modification of PCA is Nonlinear PCA described in [74]. The NLCPCA uses the architecture of the PCA network, but with complex variables (including complex weight and bias parameters). The application of NLCPCA on test problems confirms its ability to extract nonlinear features missed by the CPCA. For similar number of model parameters, the NLCPCA captures more variance of a data set than the alternative real approach (i.e. replacing each complex variable by two real variables and applying NLPCA). The NLCPCA is also used to perform nonlinear Hilbert PCA (NLHPCA) on complex real data.

In this work, we have used gradient descent based algorithms while new configurations and training algorithms have been proposed. These novel developments primarily deal with issues of **local optima** and **nonstationarity** in order to obtain better results in complex environments.

In the next section we present a few examples of NN implementations where gradient descent based learning is employed.

2.4 Implementation Example

An artificial neural network is developed to perform two main functions, pattern recognition and function approximation. In this thesis, two examples were used to test these properties of ANN. A simple load forecasting problem is adopted to test function approximation application and traffic incident detection for classification applicability of ANN. These two implementations work as learning platform for ANN.

2.4.1 Function Approximation

Function approximation is identification of the underlying mathematical model of the system by using its input signals and output signals as parameters. Thus in a way it's a black box model trying to mimic the linear or nonlinear behavior of system. Artificial Neural Networks are used for the task of function approximation on the basis of universal approximation theorem that the Neural Networks with atleast 3 layers and sufficient number of neurons in the hidden layer with appropriate nonlinear activation function (sigmoidal, Gaussian, tanh etc) is capable of assuming any nonlinear function. A classical example of function approximation is time series forecasting. In this problem the objective is to predict the future values in the time series using the past values as the input to the neural network. It is believed that the relationship between the past values and the future value in the time series follow a mathematical pattern which is derived by Neural Networks. In a typical example, to get familiar with this application of ANN, load forecasting was done using a 3 layer Network. The data used was from the California electricity market. Multilayer

Perceptron based neural network was used to approximate a time series. This time series was the hourly load variation for a zone. The 3 layer neural network (one hidden layer) was used with “n” neurons in the input layer and a single output neuron. The input layer contained load data for n consecutive hours. The output was the corresponding predicted load for the hour “n+1”. The network was trained using back propagation algorithm gradient descent algorithm. The data used for the training was the data for 4 months. The trained network was tested on next 2 month data. The neural network was observed to have a very good prediction capability because of its function approximation property. The MAPE obtained on the test data was 1.37%. The performance curve is shown in fig.2.3 in which the predicted and actual curves on the test data are shown on the same plot.

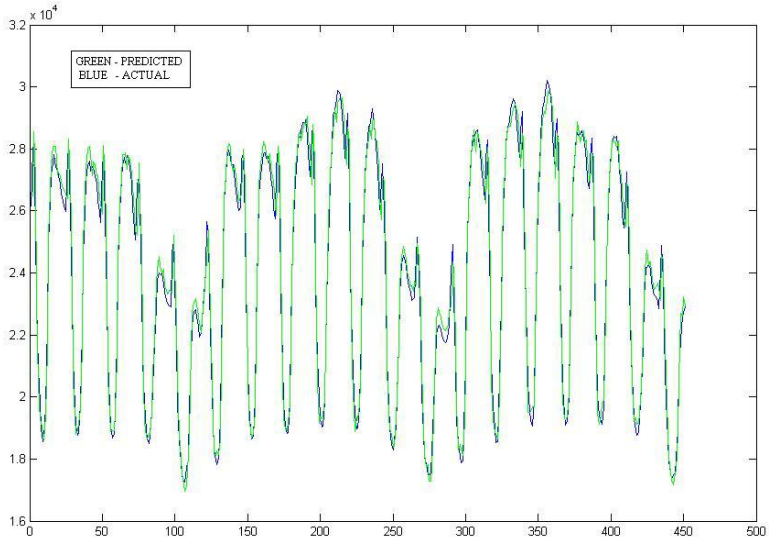


Fig 2.3 Predicted and Actual data for the Load Forecasting Problem in California Market

Then this network was tested on the next 2 months data which had similar variation nature but very high values of load. The MAPE obtained on this data was 2.4% without any

adaptation. This shows that the trained neural network has a good capability of approximation even on a completely different data with similar nature.

Pattern Classification is one of the most important applications of Neural Networks. This application was studied through another implementation of Neural Networks for freeway traffic incident which is mentioned in next section.

2.4.2 *Pattern Classification (TRAFFIC INCIDENT DETECTION ONFREEWAYS)*

Pattern Classification includes pattern and sequence recognition, novelty detection and sequential decision making. Based on the above mentioned universal approximation theorem, it can be stated that any nonlinear classification boundary in arbitrary dimension space can be formed so as to classify patterns. Since the focus of this research is neural networks, this particular work was done to understand the classification capabilities of neural networks. Classification of traffic information collected from freeways is one of the examples of two class pattern classification. The objective in this classification problem is to classify the signal from the traffic flow (such as traffic volumes, vehicle speed and traffic congestion) as input parameters and divides the source of the signal as coming from incident condition or a non incident condition. Reduced multivariate model based neural network was tested for this classification task.

Automated Incident detection (AID) is an important part of the modern traffic management system due to severe traffic congestion caused by incidents [75]. Freeways incidents are non-recurring events such as accidents, stalled vehicles, spilled loads, temporary construction and maintenance activities that disrupt the normal flow of traffic which persists for some amount of time.

Four parameters are used as the performance measure of an AID algorithm which are Detection Rate (DR), False Alarm Rate (FAR), Misclassification Rate (MCR) and Mean Time To Detect (MTTD) [76-79]. The desirable features of an ideal AID are 100% DR with a 0% FAR with minimal MTTD. The detection rate and false alarm rate of a practical AID system needs to be improved for it to become a primary source of incident detection.

Since the temporal and special disruptions exist for some time in the case of incident, this persistency has to be used for the classification task. This makes the input dimension large and thus classification task difficult.

A newly developed Reduced Multivariate (RM) based neural network has been employed for learning this classification task. RM model based neural network solves most of the intricacies associated with freeway incident detection. The localization properties of RM make it an excellent classifier. It carries out dimension reduction which makes it tractable even when the training size is large. Presence of a single column weight vector, which represents the coefficients of the polynomial model, makes the estimation task even easier. Use of appropriate Least Square estimator solves the estimation vector optimally and with fast computational speed. This property makes it particularly suitable for freeway incident detection task. Simple mathematical form makes the structure of polynomial models very easily implementable. Also there is only one weight vector between 2nd layer and 3rd layer to be estimated which allows for implementation of linear least square estimators. In work a RSVD and Gradient Descent based estimator are experimented separately. The weights correspond to the coefficients of nonlinear input components in the polynomial expression. Input patterns are formed using the training data. This input terms and its nonlinear products are directly presented to layer 2 with unity weights which are not

changed during the estimation process. The second layer consists of power, summation and product terms of the first layer terms so as to form the higher order polynomial terms. The linear combination of these terms with weight parameters is then calculated in layer 3 which gives the final output. As this is a two-category classification problem, the network is trained to give “1” as output for class 1 (incident) and “0” (non incident) for class 2.

The parameters involved in the input pattern are-

Speed - average speed from all lanes in the past 30 seconds (mph).

Occupancy – Average Occupancy of all lanes in the past 30 seconds (%).

Volume - average lane volume counted in the past 30 seconds (vphpl, vehicle per hour per lane).

The RM model showed good DR vs FAR characteristics as well which is a good characteristic of an AID. Fig. 2.4 shows the DR vs FAR curves of the RM model when implemented with Gradient Descent and RSVD based LSE as the classification threshold is varied. Both the curves follow similar characteristics with somewhat better performance in case of gradient descent based estimator. Ideally this curve should be a right angle as it should be passing through 100% DR and 0 % FAR.

TABLE 2.1
PERFORMANCE OF RM MODEL WITH GRADIENT DESCENT AS ORDER OF POLYNOMIAL IS VARIED

Order (r)	DR(%)	FAR(%)
3	89.7	1.7
4	90.4	1.2
5	91.8	0.86
6	92.5	0.46

7	93.7	0.27
8	94.0	0.18

TABLE 2.2
COMPARISON OF PERFORMANCE MEASURE OF VARIOUS ALGORITHMS

Algorithm	DR(%)	FAR(%)	MCR(%)
BP	87	0.10	5.34
Classical PSO	91.3	0.16	6.38
CPNN	91.3	0.37	4.78
CONFS	90.91	0.25	4.4
RM (with gradient descent), r=9	94.2	0.14	1.7

From the results, it can be seen that using a simple multivariate polynomial model based on a three-layer Neural Network structure is capable of solving the freeway incident detection problem quite successfully. The motivation for the use of reduced multivariate polynomial model comes from the fact that it has been particularly useful in solving pattern classification problems in the past with large number patterns available. Freeway incident detection is one such class of problem, although the number of features in this problem is comparatively large due to persistency of average lane speed, average occupancy and average volume during incident conditions. Also since this Neural Network based polynomial model takes form of linear equations, the training of the estimation parameters is simple and fast. Thus it also allows for use of other least square estimator techniques like recursive SVD based least square estimator.

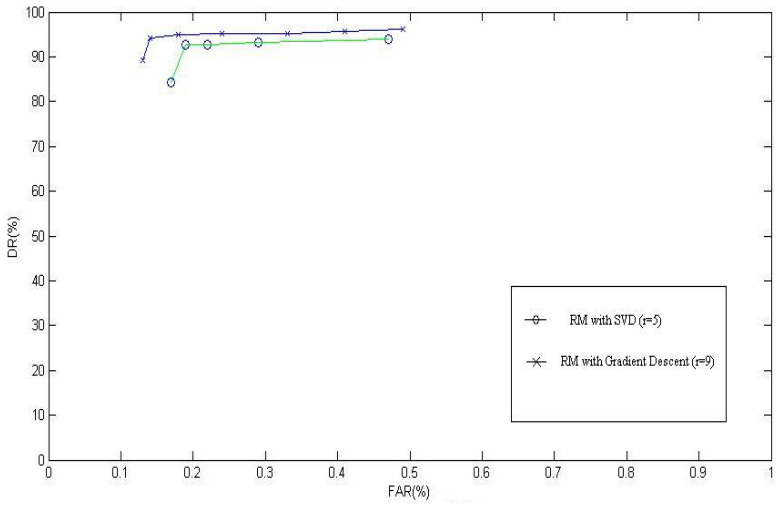


Fig. 2.4 Variation of DR and FAR as classification threshold is changed.

2.5 Summary

In this chapter a brief overview of neural network was provided in order to gain a basic understanding of NN based modeling. Various issues associated were discussed and simple implementation examples were provided. From the load forecasting problem discussed here, **local optima** is identified as the major obstacle in NN learning. The most discussed and helpful solution to this issue is to obtain heuristic information about the system and employ the obtained information to assist learning process in achieving global optimal solution. This same approach is adopted in this thesis.

The function approximation problem discussed in this thesis is electricity price forecasting in deregulated markets, which is why in next chapters, these markets are extensively studied in order to understand the factors affecting volatility of time series while in chapter 4, the dynamic characteristics of time series are studied.

CHAPTER 3

Deregulated Electricity Markets and Volatility

Markets for electric power in many countries around the world are rapidly deregulating the processes for power generation and distribution. Regulators had previously controlled the market by supply costs based fixing of prices. Currently markets represent a competitive interaction of supply and demand. Until restructuring of the electric industry, wholesale electricity markets were primarily based on bilateral contracts and cost-based power pools. Distribution utilities used to enter in cost-based, long-term contracts to meet baseload demand. As demand is varied on a short-term basis from their forecasts, distribution utilities would also enter into cost-based short-term transactions in order to match actual demand with supply. Power pools were developed to settle these short-term transactions on a variable cost basis. Restructuring of the electric industry has allowed several markets to transform cost-based bilateral contract markets or power pools into deregulated poolco markets [80-81]. Poolcos are similar to power pools as they operate in the short term, but differ from them in that the price of power is determined by market forces, not costs or regulation. In areas where poolcos are established, most power is eventually purchased through poolcos rather than on contract.

In a poolco market, generation owners send bids to the system administrator for units owned by them. These bids represent the prices at which owners are willing to sell power from specific units for a specified time period, usually the next 24 hours. The system administrator dispatches units in order of lowest to highest bid as needed to meet demand for all participants on a continuous basis. The bid price of the last unit dispatched during any

given hour is set as the market clearing price for that hour. All units dispatched during that hour have the same market clearing price regardless of their bidding price.

In a perfect poolco, generation owners bid their production costs [82]. Market power refers to the ability of one or more generation owner(s) to manipulate the market to their advantage for a sustained period of time, causing prices and profits to increase.

In a poolco, objective of generating firms is to increase the market clearing price since it is paid to all units dispatched in each time interval. There are two principal mechanisms by which players exercise market power in a poolco. The first mechanism, strategic bidding, involves players' bidding prices above the production costs of their generating units with the objective of increasing the market clearing price. The profit from "bidding up" the market clearing price can outweigh the risk of being forced on an undesirable price by a competitor. In fact, the strategy of "bidding up" the market clearing price is always more profitable than bidding marginal costs in deregulated markets.

In strategic bidding, the bids submitted by generating firms apply to the next 24-hour period. Since the demand for electricity fluctuates over any 24-hour period, players anticipate these changes in demand in their construction of a strategic bidding schedule for this period. Generating players can devise strategic bidding schedules in order to increase market clearing prices so that it exceeds the short-run marginal costs of generation in almost every hour of the day and still remain safe from being undercut by competitors. Strategic bidding also proves to be a factor in future bilateral contract markets. As owners "bid up" the price of electricity in poolcos and spot markets, they enter into future bilateral contracts if the expected profitability of those contracts is as high as what they can expect in the spot market. Strategic bidding in poolcos and spot markets has a direct impact on

bilateral contract prices. If owners in a poolco market are found to have market power, then those owners would almost certainly also have market power in a bilateral contract market.

The second mechanism for exercising market power involves players' withholding some of their capacity in the bidding process in order to create more expensive units higher up the system-wide supply curve to fix the market clearing price that would be the case otherwise.

3.1 Alternate Deregulation Models

Wholesale bids and offers in electricity are typically cleared and settled by the market operator or a special-purpose independent entity charged exclusively with that function. These operators do not clear trades but often require knowledge of the trade in order to maintain generation and load balance. The commodities within an electric market generally consist of two types: Power and Energy. Power is the metered net electrical transfer rate at any given moment and is measured in Megawatts (MW). Energy is electricity that flows through a metered point for a given period and is measured in Megawatt Hours (MWh). For most major operators, there are markets for transmission congestion and electricity derivatives, such as electricity futures and options, which are actively traded. These markets developed as a result of the restructuring of electric power systems around the world. There exist different ways in which the deregulation of market can be implemented. In a **Centralized Dispatch Model,** Retailers (either utilities or competitive suppliers) buy all of their needs from pool, resell to end users. All generators bid into pool on an hourly basis Pool dispatches generation from lowest cost bid to highest cost bid. Highest cost bid that gets dispatched becomes market clearing or "spot" price. All generators that are dispatched are paid the spot price. This type of deregulation is carried out in markets such as PJM. In

Customer Choice Model Customers may contract directly with generators or competitive suppliers (power marketers) for their own needs. Network operator runs transmission system, does planning and scheduling, balances supply and demand through bid-in balancing market, and is responsible for reliability. In **Vertically-Integrated, Incremental Wholesale Competition Model,** Utility affiliates may also bid if permitted by state regulators [83]. Over time, more and more generation is acquired through purchase power agreements, rate base diminishes. Transmission and distribution planning and operations continue to be performed by integrated utility. Integrated utility also distributes power and makes retail sales at rates set by state regulators. A **wholesale electricity market** exists when competing generators offer their electricity output to retailers. The retailers then re-price the electricity and take it to market.

3.2 Factors Affecting Volatility

Factors that influence aggregate demand among local-market distributors by pool include weather, season, and regional concentration and location of retail customers; aggregate supply is influenced by the location of generators, their market concentration, the transmission structure and the bidding and auction process. As a result, deregulated prices in these markets are characterized by volatility that varies over time and occasionally reaches extremely high levels, commonly known as “price spikes.” Understanding the volatility process is critically important to distributors, generators and market regulators as it influences the pricing of derivative contracts traded on electric power prices that allow them to better manage their financial risks, and also to analyze who attempt to model the underlying mathematics and physics behind this phenomenon.

How identifying different factors influencing the volatility process is complex and still represents an important research challenge in spite of dozens of studies that have been devoted to this question in the economics, statistics, mathematics, and engineering literatures. Common features in the electricity price volatility processes spread not only across markets but also across different major wholesale electricity trading hubs or pools within these markets. These markets are worthy of study, not only because they represent among the largest competitive markets for electric power in the world, but also because they offer interesting institutional features among the different trading hubs within the transmission networks. This is useful because what differentiates electricity is its non-storability as a commodity, which exaggerates the impact of supply and demand shocks, and the complex physical constraints, which govern the flow of power within transmission networks [12]. Factors unique to specific hubs, such as weather, the size and concentration of local generators, distributors, and the retail market, leads to locational price differences, but transmission systems within networks of these geographically-diverse pools should also lead to commonness across hubs.

We observe that chaos and time-dependent jumps are very important statistical features of price volatility across all hubs in each of the three markets we study. However, the magnitude and persistence of the volatility process and the magnitude and intensity of the jumps are varied. The next section will describe briefly the institutional features of the US, Australian and Europian markets.

3.3 Models of Spot Prices

Futures, forward and options contracts exist in a number of electricity markets and spot price dynamics are important for pricing these contracts. This is a meaningful alternative

specification, since a number of mathematical models of electricity prices propose one-factor diffusion models with constant volatility. Another mathematical approach dispenses with time-series dynamics and sets up the pricing as an linear-programming algorithm [84]. Examples of equilibrium models of pricing and hedging in forward markets include [85-88]. Ref [88] show that the forward prices for day-ahead prices in the Pennsylvania-New Jersey-Maryland (PJM) hub in the U.S. is related to price uncertainty, which they model as a GARCH process, as well as demand uncertainty and price shocks, which they model from load dynamics. Kellerhals [85] employs a stochastic-volatility model, which he operationalizes with a Kalman-filtering algorithm.

One of the key components in liberalized power sectors is the short-term electricity market, where hourly energy prices are set [4]. The market can be settled by two main settlement mechanisms, namely, the pay-as offer (also referred to as pay-as-bid) mechanism, where each selected supplier is paid at its offer price, and the pay-at-MCP mechanism, where all selected suppliers are paid at a uniform market clearing price (MCP), usually the price of the most expensive selected offer. However, in practice, the pay-at-MCP settlement mechanism is widely accepted and used for payments [5] and [6]. Companies that trade in electricity markets make extensive use of price forecast techniques either to bid or hedge against volatility. A producer that is able to predict pool prices can adjust its price/production schedule depending on hourly pool prices and its own production costs. Similarly, once a good next day price forecast is available, a large consumer can derive a plan to maximize its own utility using the electricity purchased from the pool. Besides, a good knowledge of future pool prices is very useful in valuating bilateral contracts more accurately [7]. It has been shown that MCP is most important to determine settlement costs and has significant impacts on forward transactions outside of the ISO markets. It is noted

that after the day-ahead market, where the major part of the total energy is traded, subsequent short-term market mechanisms (such as, intraday markets, ancillary reserves, and real-time markets) can be executed in order to provide the final balance between power generation and demand [6]. However, usually the major part of the volume traded in the electricity markets are related to the day-ahead market[8]. So, this paper focuses on day-ahead electricity price prediction.

In electricity markets, electrical energy is traded as a commodity. However, electricity has distinct characteristics from other commodities. The electrical energy cannot be considerably stored and the power system stability requires constant balance between generation and load. On short time scales, most users of electricity are unaware of or indifferent to its price. Transmission bottlenecks usually limit electricity transportation from one region to another. These facts enforce the extreme price volatility or even price spikes of the electricity market [6] and [9]. Besides, some researchers claimed that a uniform auction worsens spot price volatility as compared to a discriminatory auction [7]. Generation companies (GENCO) can create price volatility and outliers by their strategic biddings or, in other words, through exercising market power [1]. Since in a competitive environment all participants have the freedom to operate independently, the overall level of uncertainty in the operation of the power system increases, and the variables that might be relevant proliferate [10]. The real-time market activities can affect MCP and so the day-ahead MCP prediction is difficult under this market structure [11], Neural network based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method. Generally, when there is no transmission congestion, MCP is the only price for the entire system. However, when there is congestion, the zonal market clearing price (ZMCP) or the locational marginal price (LMP) could be employed. ZMCP may be

different for various zones, but it is the same within a zone. LMP can be different for different buses. LMP is the sum of generation marginal cost, transmission congestion cost, and cost of marginal losses, although the cost of losses is usually ignored [9]. When there is no congestion, LMP is the same as MCP. When there is congestion, the transmission line constraints are considered in order to balance supply and demand at each bus. The marginal cost of each bus is the LMP.

Market clearing price (MCP) is the lowest price that would provide enough electricity from accepted sales bids to satisfy all the accepted purchase bids. At MCP, total sales bids in their merit order would be equal to the total purchase bids down to that price in their merit order. In the presence of transmission constraints, the price of energy in constrained area, i.e., zonal market clearing price (ZMCP) or locational marginal price (LMP) is higher than the MCP [89]. LMP is defined as the price of supplying the next MW of load at a specific location, considering the generation marginal cost and delivery constraints of the physical network. ZMCP may be different for various zones, but it is the same within a zone. LMP can be different at different buses [90]. Besides, LMP forecasting is more important for market participants and it is more complex than MCP forecasting. The mainland Spain EM, a duopoly, is having a dominant player, which changes the price by strategic bidding and thus it is hard to predict the electricity prices accurately for the next day [6]. The PJM market handles congestion through LMP [91]. Publicly available data set of load weighted day-ahead MCP of the PECO control zone is considered for PJM market.

3.4 Market Design, Market Power and Pricing

Numerous studies in industrial economics show how the design of markets can influence price behavior. The attributes of markets that these studies focus on include the price elasticity of demand, concentration of ownership and capacity of generators,

generation technology, organization of pools (whether participation is voluntary or mandatory), transmission market structure and pricing, types of auctions (uniform versus discriminatory) and supply-curve bidding rules. With some exceptions, careful time-series modeling of spot prices is not emphasized in these studies. Ref [92] is an exception where author examines the design of electricity markets in England and Wales, NordPool, Australia and New Zealand and confirms that industries with a larger component of private participation in the generation market are associated with higher volatility of prices. He also shows that markets with mandatory participation in pools have higher price volatility.

Authors in ref. [93-94] offer clinical studies of the PJM market during the summer of 1999, California during the summer of 1999 and California power crisis in 2000, respectively, in which they relate price spikes to non-cooperative oligopoly behavior. These studies emphasize the important differences in the institutional features of the electricity markets around the world and their impact on pricing and volatility.

3.4.1 The Electricity Supply Industry in the U.S., Norway/Sweden and Australia

a) Australia

Prior to 1997, electricity supply in Australia was provided by vertically-integrated publicly-owned state utilities with little interstate grid connections or trade. The Australian National Electricity Market (NEM) was created in 1997 from the merger of the Victoria pool and the New South Wales (NSW) pool and the NEM Management Company (NEMMCO), a self-funding company owned by participant states, jointly manages system operations. There are now five trading regions (Australian Capital Territory, NSW, Queensland, South Australia and Victoria), all of which are interconnected. Australia runs about 80 percent coal-fired generation (15 percent renewable, hydro) and most of the consumption is in the eastern states of NSW, Victoria and Queensland.

Demand-side bidding is allowed and wholesale customers can bid into the pool. Generation of power is highly concentrated: in Victoria, there are five generators and five distribution companies, whereas in NSW, there are two generation and six distribution companies. In 2003, all electricity consumers were able to choose between electricity retailers. Bids are used to construct a “merit order” and generation is scheduled according to this merit order and regional spot prices are calculated for each five-minute period from actual supply and demand. Hourly and half-hourly prices are constructed as the average price of these five-minute prices. Supply curve bidding by generators can be in price-quantity pairs and 10 such pairs can be submitted each day. Re-bids and default-bids can be issued under certain restrictions. NEM is a mandatory auction market in which generators of 30 MW or larger compete. NEMMCO manages the wholesale electricity market and settles the short-term forward market.

It should be noted that there has been further restructuring in Australian market since 2003, however the deregulated market previous to 2003 has been studied in this work and the corresponding data has been employed

b) United States

The 1978 Public Utilities Regulatory Policies Act (PURPA) initiated U.S. deregulation from a collection of regulated, local power monopolies to a competitive market of independent energy producers and distributors [95]. There are over 3,000 electric utilities in the U.S. that deliver power to customers. Most utilities (“wired companies”) are exclusively distribution utilities, purchasing wholesale power from those that generate power and distribute it over transmission lines owned by the larger “merchant power” utilities to customers. Merchant power companies, or generators, generate electricity and trade it on the open market and control 25 percent of all U.S. power plants. About 53 percent of net electricity generation is

thermal (coal-fired) followed by nuclear (20 percent), natural gas (14 percent) and hydropower (8 percent).

The U.S. bulk power system has evolved into three major networks, or power grids: Eastern Interconnected System, Western Interconnected System and the Texas Interconnected system. Utilities within each power grid coordinate operations and buy and sell power among them. Reliability planning and coordination is conducted by the North American Electric Reliability Council (NERC) and its ten regional councils. Electricity flows over all available paths of the transmission system to reach customers. Independent System Operators (ISOs) are regionally-based non-profit entities created by state regulators to manage the transmission grid in their area and maintain reliability by making a real-time spot and day-ahead forward market in various trading hubs. The major trading hubs (with acronyms in parentheses) are Cinergy (CIN, Ohio, Indiana), California North-Path 15 (NP15), California and Oregon Border (COB), Four Corners (FC, for Utah, Colorado, New Mexico, Arizona), Palo Verde (PV, Arizona), Mead (MEAD, Nevada), Mid Columbia (MID, Washington), Entergy (Missouri), New York, New England and PJM.

In the PJM market, price spikes and the related high volatility are mostly due to the lack of generating capacity that can sometimes bring the electric system to emergency conditions. A situation of this kind took place, for example, in July 6, 1999, when an unexpectedly high relative humidity caused the load to suddenly increase due to air conditioning equipment, requiring the implementation of emergency procedures. During normal conditions, the volatility of this market is a little higher if compared with the volatility of the other markets analysed; this is more evident in summer time, which corresponds to yearly peak-loads. Volatility increases in peak load conditions are certainly due to a poor installed capacity.

c) Ontario

The Ontario electricity market is unique because of various reasons; for example, even after deregulation, about 75% of generation capacity is held by one single entity, and there exist various kinds of price and revenue caps for wholesale market participants as well as for retail customers. Moreover, Ontario is a single-settlement realtime market, unlike the other four adjacent North American electricity markets- the New York, New England, Midwest, and PJM markets- which are two-settlement ones. Finally, the Ontario power network is directly connected to the New York and Midwest electricity markets and indirectly connected to the New England and PJM markets. It is also connected to the regulated utilities in Quebec and Manitoba, both having significant energy transactions with other utilities in the United States. In view of this, the operation of the Ontario electricity market can significantly impact the North American North- East and Mid-West power interconnections, and hence its structure, operation and outcomes need close examination.

d) Spain

Spanish market shows the lowest price volatility among the markets analysed. This behaviour is mainly due to two reasons: - high amount of installed capacity; in 1999, the overall reserve capacity referred to the peak load was over 37%, while in 2000 it exceeded 32%; - high stability of fuel prices; in Spain, most of the electric energy produced during 1999 and 2000 came from nuclear (35%) and coal fired (43%) power plants (it must be taken into account that national coal is economically boosted by the Spanish government). In addition, any possible collusive behaviour among generating companies (the two main companies have a market share of over 70%), that could increase spot market prices and volatility, has been limited by the enforcement of a stranded costs refund system: the higher the market price, the lower the stranded costs refund.

3.5 Summary

In this chapter, the structure of deregulated electricity market has been briefly discussed and the price formation mechanism has been studied. Various deregulated markets worldwide employ different deregulation models, which is why the volatility behaviour differs. However all these markets share one same attribute, which is, occurrence of spikes at irregular intervals with chaotic motion in the interspike interval. In the next chapter, these dynamics of electricity price time series have been studied. The invariant measures of time series, which are the inherent chaotic exponents have been calculated and employed for differentiating different markets.

CHAPTER 4

Dynamic Characteristics of Electricity Price Time Series

Electricity price time series are the class of complex time series with chaotic features. It is evident from fig. 4.1 and 4.2 that no clear information can be extracted directly and no predictive patterns can be seen from electricity price time series, its projected phase space or the corresponding recurrence plot. Chaotic systems have strange attractors characterized by properties of the attractor independent of particular trajectory which are called invariants of the system. The study of dynamic characteristics of this kind of time series include study of invariant sets of time series and, for this particular work, extraction of dynamic attributes which are the key to understanding and modelling of neural networks based on time series.

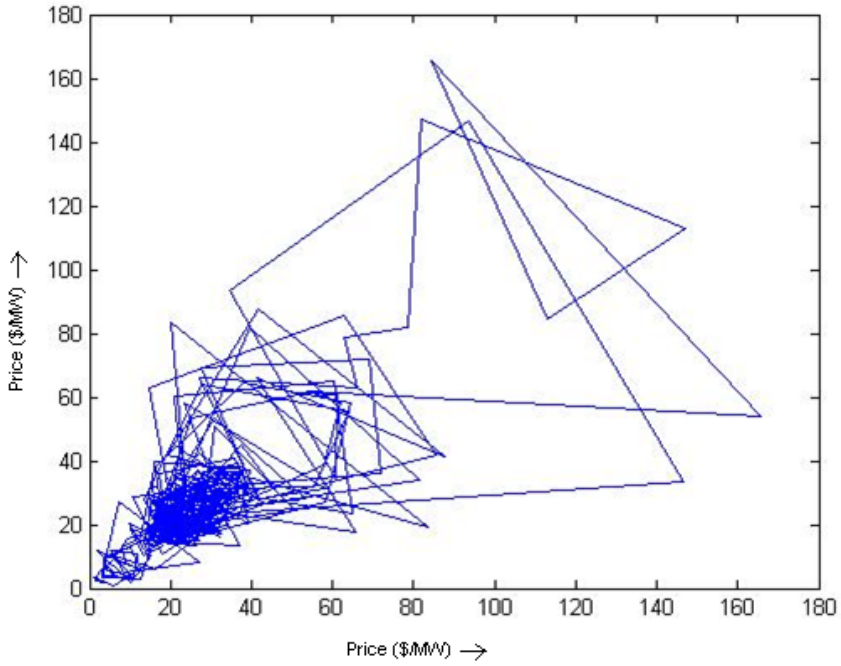


Fig.4.1. 2D Projection of phase space of Ontario time series

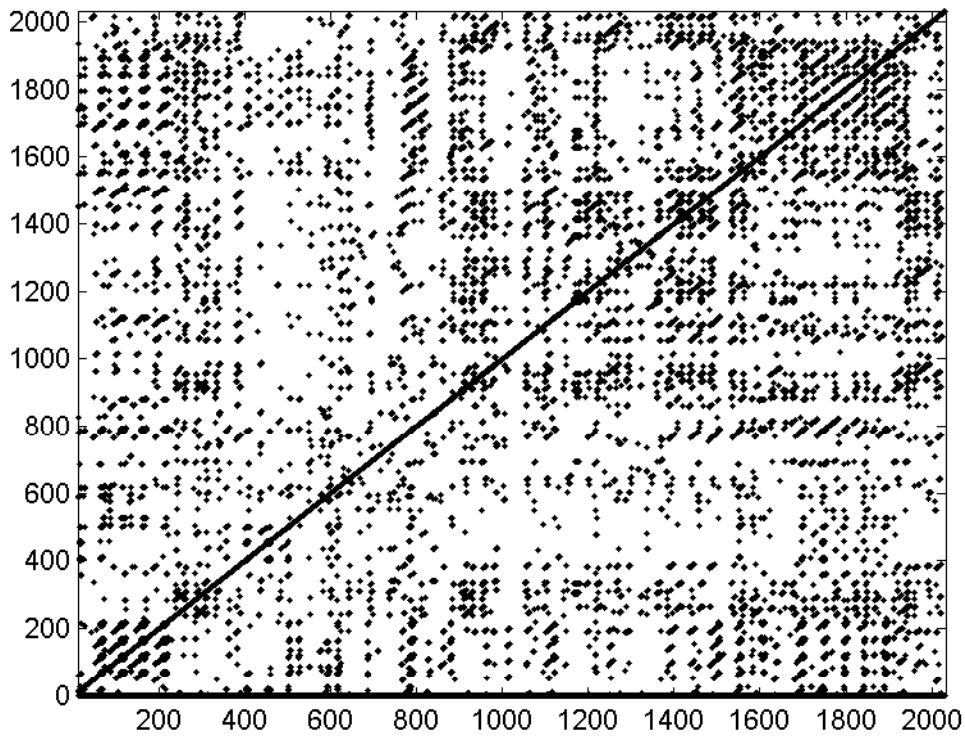


Fig 4.2 Recurrence Plot [96] of Ontario time series

The invariant set of dynamical system is a general entity in nonlinear dynamics. The important stable invariant sets are attractor in phase space and its attributes which are associated with asymptotically stable motion. Unstable invariant sets affect the global dynamical behaviours of dynamical systems and exhibit more complex properties. In this work we study both stable and unstable invariants of time series which include, embedding dimension, largest Lyapunov exponent, finite size Lyapunov exponent, scale dependent Lyapunov exponent and fixed point dynamics. The details are given in sections 4.1-4.5.

4.1 Embedding Dimension

According to Takens' delay embedding theorem [97], a time series in phase space can be geometrically fully represented by set of points in phase space of dimension *n* provided that *n* is sufficiently large to be able to uniquely locate the position of each pattern point in phase space. Such an embedding space preserves the invariant characteristics of original attractor.

One of the unique characteristics of chaotic system is the non integer value dimension of original phase space containing the evolution of dynamical system, and such chaotic systems are also known as fractals. Standard manifolds in Euclidean geometrical phase space are not fractals but have integer dimensions D = d. The primary importance of fractals in dynamics is that strange attractors are fractals and their fractal dimension D is simply related to the embedding dimension of dynamical variables needed to model the dynamics of the strange attractor. The first step towards calculating embedding dimension is to measure fractal dimension of given dynamic system. In this work, method of box counting dimension is employed where the trajectory in phase space where the measurement set is covered with small hypercubes of dimension *r* [98]. Let $M(r)$ be the number of such hyperspheres that contain the trajectory. The dimension is then defined as

$$D = \lim_{r \to 0} \frac{\partial \log \sum M(r)}{\partial \log(r)}$$

In practical, the limit $r = 0$ is not feasible. Instead, the number $M(r)$ is measured for a range of small values of r and the dimension D is estimated as the slope of straight line log(M(r) versus log(r) plot as shown in Fig 4.3. Takens' embedding theorem states that if a time series belongs to attractor represented by a D-dimensional manifold then the topological properties of this time series are equivalent to properties of embedding formed by vectors of dimension n>2D+1.

Applying the theory shown above, the embedding dimension and fractal dimension of price time series are obtained. The nature of values of these measures is shown in Table 5.1 below. Given the suspected nonstationarity in the spiking regime of time series, only the non spiking part of phase space is used for these calculations

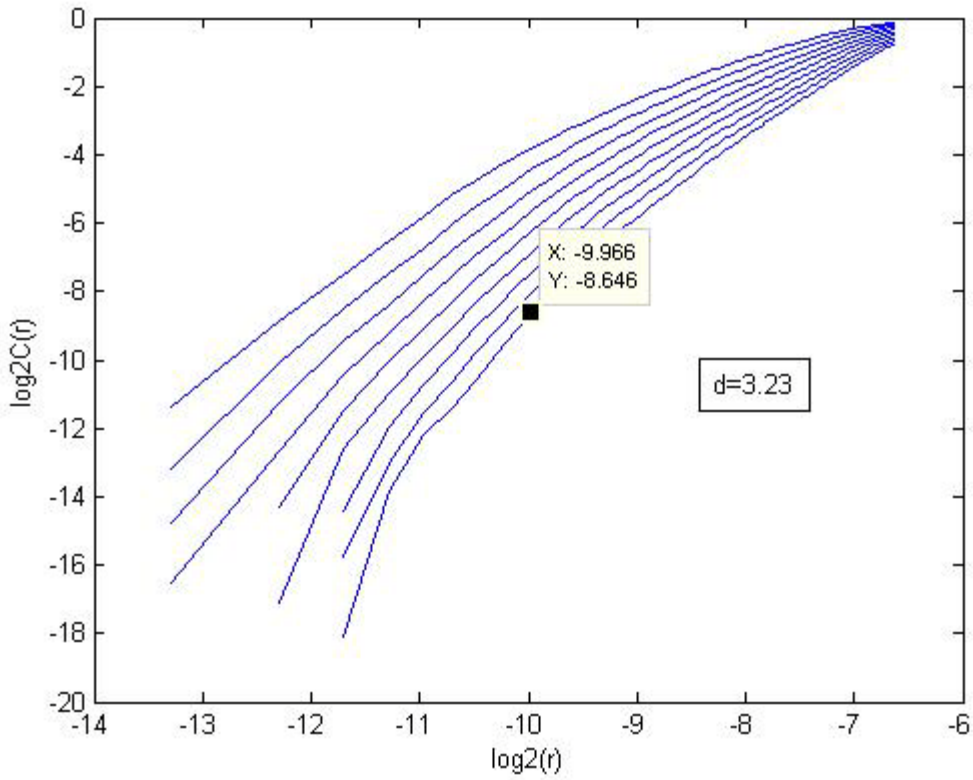


Fig.4.3 Fractal Dimension plots for California market in the non spiking region of time series for embedding dimensions 4 to 22 are shown for even numbers. Fractal dimension is estimated as slope.

Table 4.1
Embedding Dimension Calculation for Various Deregulated Markets

Market	**Fractal Dimension (D)**	**Embedding dimension (n)**
PJM	3.34	8

Ontario	4.73	11
Victoria	3.87	9
NSW	3.58	9
Spain	2.9	7
California	2.43	6

From the table 4.1 it is clear that time series corresponding to each deregulated market belong to class of chaotic system with fractal dimension. The above analysis shows that time series in Ontario and Australian markets belong to phase space with largest dimension. The value of embedding dimension is estimated which is used in later stages for time series modelling and reconstruction of phase space.

4.2 Fixed Point Characteristics

Location of fixed points and the dynamics in their neighbourhood are the inherent invariant features of time series which are responsible for local and global dynamics in phase space. Most of the real world time series consist of saddle points in phase space giving rise to complicated homoclinic or heteroclinic scenarios [99]. Eigenvalues of these saddles determine the behavior of the system.

In the first step of identifying these eigenvalues and their nature, a method is adopted for estimating the position of fixed saddle point in the time delay based reconstructed phase space. Then a combined method of radial basis function and linear modelling is used for estimation of gradient and eigenvalue of linearized neighbourhood of saddle.

4.2.1 Locating Fixed Point

In the reconstructed phase space, the fixed point lies along the diagonal of embedding space. However the actual time series in which the trajectory passes through the vicinity of saddle, it is asymptotic to the spanning vectors of embedding space which are the set of straight lines corresponding to eigenvectors of fixed point. Since the trajectory doesn't pass through exact fixed saddle of phase space, the point of trajectory closest to diagonal line is used as the first approximation of saddle, which is denoted as y. Now we choose a hypersphere of radius e and form a matrix B_e whose rows constitute vectors $x_t - y$ for all the points x_t such that $\|x_t - y\| < e$. Then the spanning eigenspace of B_e is given by the eigenvectors obtained via SVD of matrix. Let the number of dimensions be d, the intersection of the hyperspace spanned by d eigenvectors with the diagonal of embedding space is the closer estimate of location of saddle. In order to obtain this location of saddle fixed point, assume the vector normal to the hyperplane as c_d, then the fixed point satisfies [99]

$$(x.u - y).c_d = 0 \tag{4.1}$$

which implies

$$x = \frac{y.c_d}{u.c_d} \tag{4.2}$$

where $u = (1 \quad 1 \quad . \quad . \quad 1)^T$.

4.2.2 Dynamics in Neighbourhood of Fixed Point

It is assumed that a nonlinear function σ describes the evolution of vector x_n to the next vector in trajectory x_{n+1}, in the neighbourhood of saddle fixed point. The neighbourhood is considered to be a hypersphere of radius e which is formed in the same manner as described above. Let the neighbourhood contains *N* points. The mapping function σ is considered to be a combination of a linear and nonlinear basis functions.

$$\sigma_j(x_n) = b_0 + \sum_{l=1}^{p} w_{lj} b_l(x_n) \tag{4.3}$$

and the weights are chosen such that the mapping error function over B_e i.e.

$$J = \sum_{n=1}^{N} \left\| x_{n+1} - \sigma(x_n) \right\|^2, \tag{4.4}$$

is minimized. This can be achieved using a Least Square solution technique. In this work gradient descent algorithm is employed for obtaining the optimal weight values. The eigenvalues near the saddle is obtained by calculating Jacobian of the obtained model.

The results are given in table 4.2 below

Table 4.2
Eigenspectrum of Fixed Point of Time Series

Market	**D**	**Eigen spectrum**
PJM	6	$(0.13, -0.09 \pm j0.25, -0.07 \pm j0.32, -0.38,)$
Ontario	7	$(0.82, 0.13 \pm j0.25, 0.02 \pm j0.4, -0.27 \pm j0.39)$

Victoria	6	$(0.433, 0.09, +0.02 \pm j0.38, -0.17 \pm j0.24)$
NSW	6	$(0.59, +0.09 \pm j1.1, -0.23 \pm j0.82, -0.27)$
Spain	5	$(0.12, -0.1 \pm j0.59, -0.22 \pm j0.27)$
California	4	$(0.07, -0.02, -0.32 \pm j0.61)$

4.3 Lyapunov Exponents

Largest Lyapunov exponent is one of the characteristics of time series which helps in detecting the presence of chaos in dynamical system and quantifying it. The assumption of determinism in the chaotic region of electricity price time series is characterized by small but positive largest Lyapunov exponent (LLE) of time series. This exponent quantifies exponential divergence of close trajectories in phase space and thus estimates the amount of chaos in the system. There are many methods available in literature for accurate calculation of largest Lyapunov exponent. In this work the largest Lyapunov exponent is obtained using fast and easy to implement method which is robust to small variations in time series and also works accurately for small size data set in presence of noise [100]. The results are shown in Table 4.3 below. Again the calculations are constrained to nonspiking region.

Table 4.3
Lyapunov Exponent for Various Deregulated Markets

Market	PJM	Ontario	Victoria	NSW	Spain	California
Lyapunov Exponent	0.24	0.69	0.5	0.73	0.17	0.09

The results demonstrate the amount of chaos in all deregulated markets studied in this work. It can be observed that Ontario and Australian markets exhibit highest amount of chaos with largest values of LLE.

4.4 Finite Time Lyapunov Exponent Analysis and Local Instability

While asymptotic properties exhibited by dynamical systems are extensively studied and provide comprehensive information about the system, properties exhibited over finite time intervals are difficult for systematic description. However, investigating transient behaviours of dynamical systems is quite important in real world systems such as electricity price time series. A typical time series plot of electricity prices exhibits transient behaviour such as intermittent short lived spikes. Lyapunov exponents are useful tools applied for measuring such sensitivity in the case of asymptotic chaos, therefore it is important to observe and analyze finite-time Lyapunov exponents for understanding and characterizing the class of transient chaos present in the system. Studies of such systems with chaotic transients conclude that the behaviour can be either chaotic or periodic. Among various tools that may be used to distinguish between chaotic and nonchaotic evolution, finite time lyapunov exponents can be used.

The key reason behind investigating FTLEs is the detection of loss of hyperbolicity in the system. Loss of hyperbolicity corresponds to loss of local stability along the direction normal to invariant manifold of time series. This loss of local stability is characterized by the change in real axis behaviour of one of the negative lyapunov exponent, or in other word change in sign of the corresponding lyapunov exponent, which adds an expanding direction of the invariant manifold. This phenomenon is popularly known as Unstable Dimension Variability (UDV). This is an important investigation as presence of UDV can severely limit the scope of predictability and determinism in time series system [101].

While it is possible to obtain FTLEs using statistical methods [102] similar to those used for Lyapunov exponents, the approach adopted in this research is that the jacobian approximation be calculated in a mathematical manner. Here a model based approach has been employed for calculation of FTLEs, reproducing correctly variation in time of the finite-time lyapunov exponents corresponding to transient chaos. A nonparametric model (RBF) based approach for calculation of FTLE has been implemented, because of excellent approximation abilities of RBF. They are defined in local neighbourhood of a point phase space with sufficient neighbours. For a system such as electricity price time series discussed herein, a numerical approximation for the Jacobian is readily available via the method described for obtaining fixed point behaviour in Section 4.2.2.

The results obtained for FTLEs for deregulated markets did not exhibit any behaviour supporting loss of hyperbolicity along the time series. The eigenspectrum of time series corresponding to deregulated markets were determined along the trajectory and it was observed that FTLEs do not change sign. One result with minor importance was obtained for 3rd largest finite time lyapunov exponent for Ontario market (the first and second largest

FTLE's remain positive during the complete time series). The result obtained is shown in Fig.4.4 where it is evident that FTLE doesn't change sign during the course of time series except at 3 locations for only one hour which is an inconclusive observation from hyperbolicity point view.

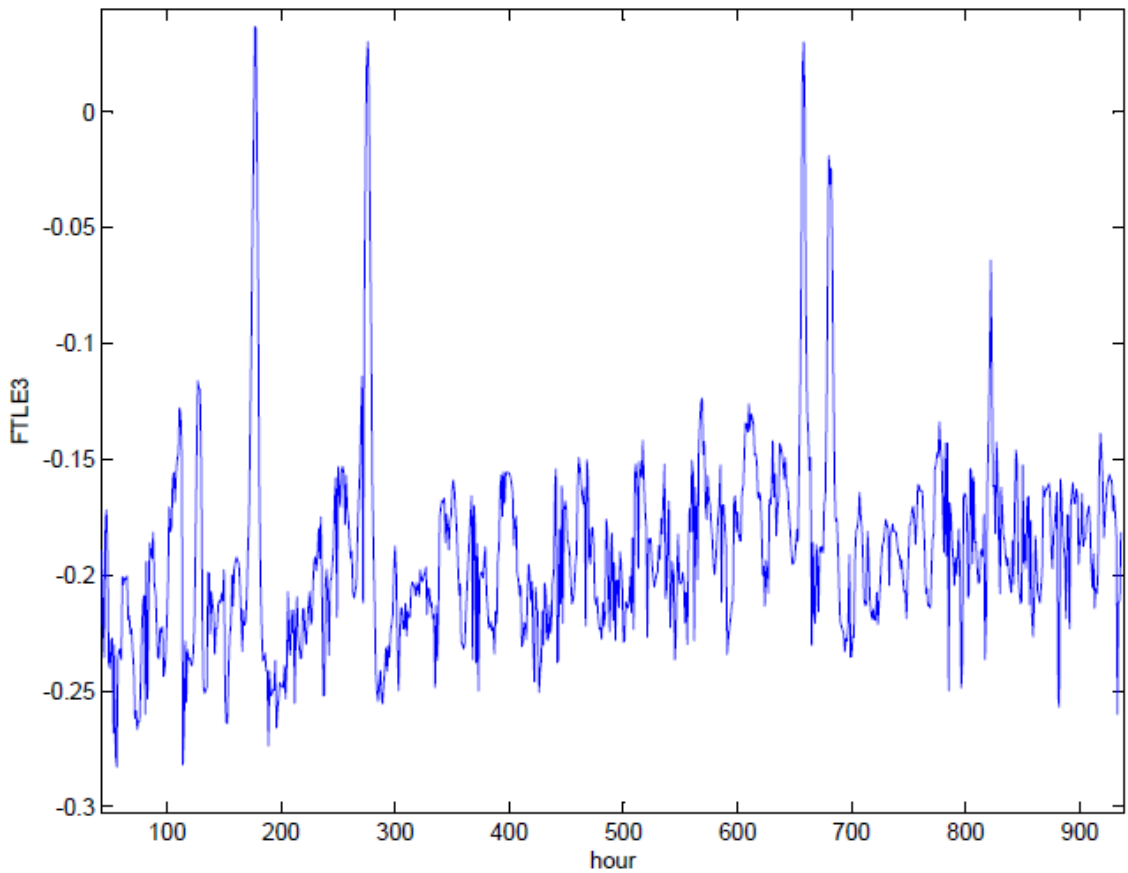


Fig. 4.4 Third Largest FTLE for Ontario Market Time Series

From the above analysis it is evident that FTLEs for electricity price markets do not lose hyperbolicity and it can be conclusively said that UDV is not the responsible mechanism behind spiking dynamics.

4.5 Scale Dependent Lyapunov Exponent (SDLE):

Despite extensive studies on electricity price time series using chaos theory, [103-106] fractal scaling analysis, and many other methods in the last decade, the issue of whether electricity price time series is chaotic or stochastic remains highly controversial. The debate can hardly be settled if one does not go beyond the standard theories of chaos and random fractals. Chaos theory is mainly concerned with apparently irregular behaviors in a complex system that are generated by nonlinear deterministic interactions with only a few degrees of freedom, where noise or intrinsic randomness does not play an important role, while random fractal theory assumes that the dynamics of the system are inherently random. To shed new light on the problem, here we employ a newly developed multiscale complexity measure, the SDLE. SDLE is calculated in this work using the method proposed in [107].

In the case of a scalar time series $x(1),x(2),\ldots,x(n)$, in the time delay based reconstructed phase space, SDLE can be obtained using divergence of distance based approach. Let us denote the initial distance between two nearby trajectories by ε_0 and their *average* distances at time t and $t+\Delta t$, respectively, by ε_t and $\varepsilon_{t+\Delta t}$, where Δt is small. The SDLE $\lambda(\varepsilon_t)$ is defined by [107]

$$\lambda(\varepsilon_t) = \frac{\ln(\varepsilon_{t+\Delta t}) - \ln(\varepsilon_t)}{\Delta t} \tag{4.5}$$

To compute SDLE, we can start from an arbitrary number of shells,

$$\varepsilon_k \le \left\| V_i - V_j \right\| \le \varepsilon_k + \Delta\varepsilon_k, \qquad k = 1,2,\ldots, \tag{4.6}$$

where V_i,V_j are reconstructed vectors and ε_k (the radius of the shell) and $\Delta\varepsilon_k$ (the width of the shell) are arbitrarily chosen small distances ($\Delta\varepsilon_k$ is not necessarily a constant). Then we

monitor the evolution of all pairs of points (V_i,V_j) within a shell and take average. Equation (4.5) can now be written as

$$\lambda(\varepsilon_t) = \frac{\left\langle \ln\left\| V_{i+t+\Delta t} - V_{j+t+\Delta t} \right\| - \ln\left\| V_{i+t} - V_{j+t} \right\| \right\rangle}{\Delta t} \tag{4.7}$$

where t and Δt are integers in unit of the sampling time and the angle brackets denote average within a shell.

The initial set of shells for computing SDLE serves as initial values of the scales; through evolution of the dynamics, they will automatically converge to the range of inherent scales. This is emphasized by the subscript t in ε when the scales become inherent, index t can then be dropped. Using this evaluation, the deterministic nature of electricity price time series can be established using the following three cases, which distinguish between low dimensional chaos, high dimensional chaos and stochastic noise.

a) For clean chaos on small scales and noisy chaos with weak noise on intermediate scales, eq (4.7) follows the behavior

$$\lambda(\varepsilon) = \lambda_1 \tag{4.8}$$

Such a definition of chaos is able to detect chaos in intermittent time series with a long quiescent phase during which neighbouring trajectories do not diverge.

b) For variation in chaos [108-109] on small scales,

$$\lambda(\varepsilon) = -\gamma \ln \varepsilon, \tag{4.9}$$

where γ>0 is a parameter.

c) For random $1/f^{2H+1}$ processes, where $0<H<1$ is called the Hurst parameter which characterizes the correlation structure of the process: depending on whether H is smaller than, equal to, or larger than 1/2, the process is said to have antipersistent, short-range, or persistent long-range correlations,

$$\lambda(\varepsilon) \approx \varepsilon^{-1/H} \tag{4.10}$$

Note that the standard Brownian motion corresponds to H = 1/2 and generally H<1/2 for electricity price time series

The result of SDLE procedure is shown in Fig. 4.5 as an example for the most volatile Ontario market. The fig indicates that curve most closely follows eq (4.9)

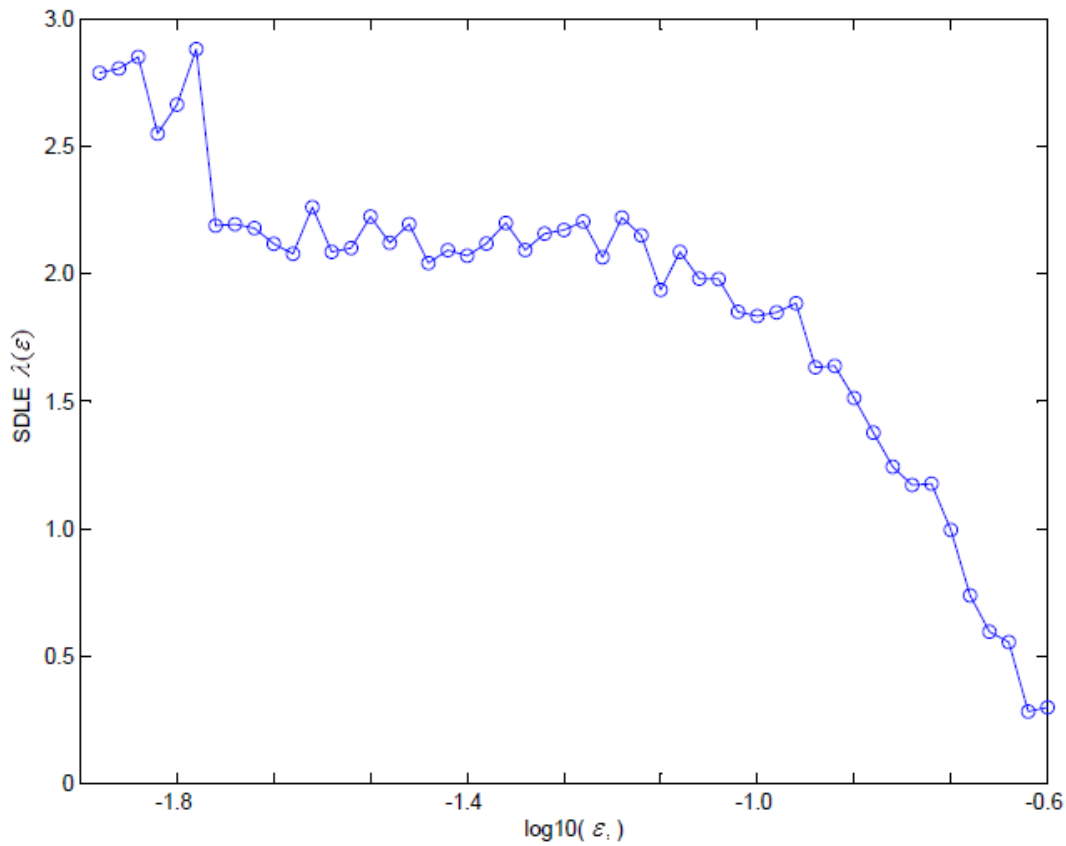


Fig.4.5. SDLE plot for Ontario Market

The above plot immediately disregards presence of stochastic fluctuations in time series while variation in chaos at small scale, hence multi-scale behavior is evident.

4.6 Summary

In this chapter some selected nonlinear measures of time series are measured and observed. The main contributions and conclusions of this chapter come from the FSLE and SDLE analysis, which are local exponents. The observation of FSLE confirms that loss of hyperbolicity, or UDV is not the underlying mechanism behind spiking. Moreover the SDLE analysis confirms that noise does not play any role behind spiking dynamics, while it is the multiscaling characteristic of time series which are responsible behind complex behavior. Largest Lyapunov exponent and correlation dimension are the key invariant measures of a

chaotic time series. The values of obtained largest lyapunov exponent and fractal correlation dimension indeed confirmed the chaotic nature of time series in the interspiking region. Moreover the correlation dimension assists in obtaining the dimension of phase space in which a phase space can be reconstructed with topological sufficiency. This embedding dimension is used later for modeling of recurrent neural networks. Another important invariant measure obtained in this chapter is fixed point dynamics. The location of fixed point and the dynamics in the neighborhood are the invariant information which ought to be an essential component of the describing dynamic model. In the next chapter, these calculated invariant measures of time series are employed for developing neural network based models. The basic motive behind this approach is to bring the developed model close to actual dynamics of time series.

Chapter 5

Electricity Price Time Series Prediction Using RNN Trained Using Invariant Dynamics

Training recurrent neural networks (RNN) is one of the key challenges in nonlinear optimization. In this chapter, an approach based on the consideration that nonlinear time series are chaotic signals has been presented for intelligent learning of Recurrent Neural Networks. A desirable characteristic of RNN training algorithms is to be able to learn the inherent dynamic of the system to be approximated. In this chapter, some of the invariant dynamic features of time series have been extracted and incorporated in the training of RNN as constraints. It is demonstrated that this approach brings the trained model closer to exact nonlinear system. Time series from modern deregulated market electricity price have been modelled using the proposed approach. After extensive comparison with benchmarks, it has been demonstrated that the results are improved considerably.

5.1 Introduction

Currently there is an exceeding interest in recurrent neural networks learning algorithms. Training RNN weights is one of the key challenges due to sensitivity of network behaviour on feedforward and feedback weights. Concerning function approximation, specifically time series forecasting, there are many issues one of which involves ability to capture the exact nonlinear features of dynamic system represented by time series. This becomes absolutely “necessary” when the system exhibits critical transitions. Most of the real world physical system have critical thresholds, also known as tipping points, at which the system abruptly shifts its state from one to another. This phenomenon is quite often seen in medicine, weather, finance etc. In earth, abrupt shifts in ocean currents may cause

climate changes. Spontaneous changes known as epileptic seizures can occur in brain EEG signals. Although it is of utmost importance to predict these changes, the prediction becomes notably hard because the system shows unnoticeable change before tipping point is reached. Moreover the prediction models are not accurate enough and thus do not hold the essential dynamic features of system to forecast reliably where the critical threshold is present in phase space. However the behaviour of dynamic system near a critical point exhibits generic properties which explain that sharp transitions in a range of complex systems are related. In phase space, critical transitions correspond to bifurcation. There exist leading indicators that occur in the non equilibrium state of dynamic system before critical transition which are invariant variables or features. Such indicators, if existent in predictor models, can dramatically improve forecasting of critical transitions as well as overall time series modelling. These indicators are the footprints of critical transition on the phase space trajectory. We try to model an RNN using a novel learning algorithm which can most accurately inherit the invariant dynamics of nonlinear system so that these indicators can be captured.

A gradient descent based training algorithm has been proposed which incorporates the invariant features of the dynamic system. This approach presents the possibility to find the optimal value of weights and biases for RNN in a gradient based supervised manner while using the invariant features as heuristic information. This is a unique attempt to employ heuristics in real world time series modelling.

For accurate dynamic modelling in supervised learning, trained model should posses the same dynamic properties as the teacher signal. The learning problem includes modifying the weight of the network so that the trajectory has specified characteristics. However, this

training is a challenging task due to complexity of the error surface where the network often converges to undesired local optima. A possible solution to this problem is intelligent weight initialization. Invariant measures, in particular fixed point dynamics can be employed as necessary heuristic information for weight initialization. In the later stages, the fixed point dynamics are enforced on the RNN model during training using projection gradient descent learning and finite time Jacobian information are incorporated using penalty function based approach. This chapter includes extensive computation of various steps involved in this process. Section 5.2 discusses RNN weight initialization employing the fixed point dynamics derived in Section 4.2. In Section 5.3, the training algorithm based on projection descent learning is derived for enforcing fixed point conditions during each learning step. In Section 5.4 the learning of local jacobian is incorporated using penalty function based approach. Finally the trained model is tested on electricity price time series data and results are presented along with discussion.

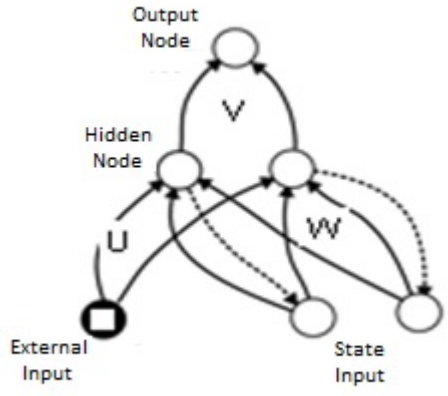


Fig 5.1 Elman State Feedback Recurrent Neural Network

5.2 Weight Initialization

The objective is to initialize U (feedforward weights from input to hidden layer), W (feedback weights) and V (feedforward weights from hidden layer to output layer). In this work the weights are initialized using the fixed point information of the system. Fixed points and the dynamics in their neighbourhood are the most important invariant information of the dynamic system as they affect the local system properties and, in most cases, global dynamics of the system. Most of the real world systems exhibiting complex chaotic properties consist of fixed points in the form of saddle and thus the phase space has homoclinic or heteroclinic orbits. Moreover it will be evident from discussion in Chapter 7 that dynamic system such as electricity price time series with intermittent spiking dynamics consist of saddle in phase space where homoclinic orbits are characterized by two time scales. The position of saddle and the eigenvalues of linearized neighbourhood of saddle are key invariant information of the given dynamic system which we enforce during weight initialization and learning of recurrent neural network.

Insights from control theory have helped in developing gradient descent based algorithms for learning of RNN. Different methods have been developed which calculate the gradient in a different and efficient manner. However the computational complexity is the main shortcoming of these approaches. So far gradient based approach has been used for learning. Most of the gradient based algorithms such as backpropagation through time [110], forward propagation [111], etc are just variation of a unified approach. In this thesis, a gradient descent based approach has been modified via incorporating invariant dynamics in the learning procedure.

5.2.1 Identifying Fixed Point Location and Neighbourhood Dynamics

In this section, a method is adopted for estimating the position of fixed saddle point in the time delay based reconstructed phase space. Then a model which is a combination of radial basis function and linear modelling is used for estimation of gradient and eigenvalue of linearized neighbourhood of saddle.

In the reconstructed phase space, the fixed point lies along the diagonal of embedding space. However the actual time series in which the trajectory passes through the vicinity of saddle, it is asymptotic to the spanning vectors of embedding space which are the set of straight lines corresponding to eigenvectors of fixed point. Since the trajectory doesn't pass through exact fixed saddle of phase space, data point of trajectory closest to diagonal line is used as the first approximation of fixed point, which is denoted as Y_0. Now we choose a hypersphere of radius ε and form a matrix $B\varepsilon$ whose rows constitute vectors $X_n - Y_0$ for all the points X_n such that $\|X_n - Y_0\| < \varepsilon$. Then the eigenspace spanning $B\varepsilon$ is given by the eigenvectors obtained from SVD of matrix. The exact number of significant dimensions of spanning space is obtained as the number of significant singular values. Let this number of dimensions be d, the intersection of the hyperspace or hyperplane spanned by d eigenvectors with the diagonal of embedding space is the closer estimate of location of saddle. In order to obtain this location of saddle fixed point, assume the vector normal to the hyperplane as c_d, then the fixed point $x_0.u$ satisfies.

$$(x_0.u - Y_0).c_d = 0 \tag{5.1}$$

or,

$$x_0 = \frac{Y_0 . c_d}{u . c_d} \tag{5.2}$$

where $u = (1 \quad 1 \quad . \quad . \quad 1)^T$.

The neighbourhood is considered to be a hypersphere of radius ε which is formed in the same manner as described above, however this time we use $x_0 . u$ as the center of hypersphere. Let the neighbourhood contains N points. It is assumed that a nonlinear function σ describes the evolution of vector X_i to the next **point** in time series X_{i+1}, in the neighbourhood of saddle fixed point. The mapping function σ is considered to be a combination of a linear and nonlinear basis functions.

$$\bar{x}_{i+1} = \sigma(X_i) = b_0 + \sum_{l=1}^{p} w_{li} b_l (X_i - X_l) \tag{5.3}$$

Here b_0 is the linear part. And the weights w_{li}, $l = 1,.., p$ are chosen such that the mapping error function over $B\varepsilon$, i.e.

$$J = \sum_{n=1}^{N} \left\| x_{i+1} - \sigma(X_i) \right\|^2, \tag{5.4}$$

is minimized. This can be achieved using a Least Square solution technique. In this work gradient descent algorithm is employed for obtaining the optimal weight values. The eigenvalues near the saddle are obtained by calculating jacobian of the obtained model. In a delay embedded phase space, the jacobian at a particular point X_i in trajectory is given by

$$\begin{bmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \\ \lambda_1 & . & . & . & \lambda_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \\ \frac{\partial \overline{x}_{i+1}}{\partial x}_{1x=x_i} & & . & . & \frac{\partial \overline{x}_{i+1}}{\partial x}_{nx=x_i} \end{bmatrix} \tag{5.5}$$

Here $\overline{x}_{i+1}$ is the output of radial basis function eq. (5.3). Lyapunov exponent of the linearized neighbourhood of X_i is given by the eigenvalues of this jacobian. The result was demonstrated and discussed earlier in table 4.2 in section 4.2.

5.2.2 Fixed point based initialization

The objective is to initialize feedback, feedforward and third layer weights using the fixed point information. Let us denote these weight matrices as *W, U* and *V* respectively. Here y_R, *X* and *z* are the scalar output, input vector and state vector respectively.

$$y_R = V \cdot f(U \cdot X + W \cdot z) \tag{5.6}$$

Here y_R is the RNN output, *X* is the input vector including bias and *z* is the state vector. The RNN following the saddle fixed point x_0 and the dynamics has to follow these three properties.

1. For input vector x_0.u, the output is equal to x_0. Or

   $$y_R = V \cdot f(U \cdot x_0 u + W \cdot z) = x_0 \tag{5.7}$$

   where y_R is the output of RNN

2. The jacobian of RNN is same as the calculated gradient of time series at the saddle fixed point. In a Takens' delay embedding based restructuring, this condition can be formulated as

$$\begin{bmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \\ \frac{\partial y_R}{\partial x}_{1x=x_0} & . & . & . & \frac{\partial y_R}{\partial x_n}_{x=x_0} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \\ \lambda_1 & . & . & . & \lambda_n \end{bmatrix} \tag{5.8}$$

here matrix on right hand side is the obtained jacobian from time series.

3. Matrices *W* should be such that state vector is an asymptotically stable equilibrium when input vector is at the saddle. This condition can be formulated for the stability of state vectors for given input as follows.

In the RNN formulation (5.6), the state evolution under constant input can be represented in continuous form as

$$\frac{dz}{dt} = -z + f(U \cdot x + W \cdot z) \tag{5.9}$$

For the given input at fixed point, it is desirable that the after short transient period, the network state reaches a steady state irrespective of the initial state. This implies that state should be attracted to a unique global equilibrium which is conditional on matrices W and probably U . These conditions are presented as a theorem.

Theorem 5.1: A network represented by (5.9) reaches a unique equilibrium state if the following condition is satisfied (where *f* is the nonlinear sigmoid function)

$$\sum_i \sum_j w_{ij}^2 < 1/\max\left(f'\right)^2 \tag{5.10}$$

Let $z_1(t)$ and $z_2(t)$ be the two solutions of (5.9). Let

$$J(t) = \left\| z_1(t) - z_2(t) \right\|^2 \tag{5.11}$$

Differentiating J with respect to time,

$$\frac{dJ(t)}{dt} = 2(z_1(t) - z_2(t))^T \left(\frac{dz_1(t)}{dt} - \frac{dz_2(t)}{dt} \right) \tag{5.12}$$

Using (5.9), we obtain,

$$\frac{dJ(t)}{dt} = -2\left\| z_1(t) - z_2(t) \right\|^2 + 2\left(z_1(t) - z_2(t)\right)^T \cdot \left[f(U \cdot x + W \cdot z_1) - f(U \cdot x + W \cdot z_2) \right] \tag{5.13}$$

Using Schwarz's inequality [112], the expression becomes

$$\frac{dJ(t)}{dt} \le -2\left\| z_1(t) - z_2(t) \right\|^2 + 2\left\| z_1(t) - z_2(t) \right\|^T \cdot \left\| f(U \cdot x + W \cdot z_1) - f(U \cdot x + W \cdot z_2) \right\|$$
(5.14)

Now by mean value theorem, for maximum possible value of *f,* the equation becomes

$$\frac{dJ(t)}{dt} \le -2\left\| z_1(t) - z_2(t) \right\|^2 + 2\left\| z_1(t) - z_2(t) \right\|^T \cdot (\max(|f'|)\left\| W \cdot (z_1(t) - z_2(t)) \right\| \tag{5.15}$$

Again applying mean value theorem,

$$\frac{dJ(t)}{dt} \le -2\|z_1(t) - z_2(t)\|^2 \cdot (1 - (\max(|f'|) \cdot \|W\|) \text{, or} \tag{5.16}$$

$$\frac{dJ(t)}{dt} \le -2J \cdot \left(1 - (\max(|f'|) \cdot \sqrt{\sum_i \sum_j w_{ij}^2}\right) \tag{5.17}$$

Let $\left(1 - (\max(|f'|) \cdot \sqrt{\sum_i \sum_j w_{ij}^2}\right) = \beta$, then multiplying both sides of (5.17) by $\exp(\beta t)$, we obtain

$$\frac{d}{dt}\left(J(t)e^{\beta t}\right) \le 0 \text{ or}$$

$$J(t) \le J(0)e^{-\beta t} \tag{5.18}$$

Given that $J(t)$ is non-negative, it follows that $J(t)$ approaches zero as $t \to \infty$, which is why starting from any initial condition, the states approach a unique equilibrium, which completes our proof.

These conditions are satisfied progressively in iterative manner. Initially we start with randomly chosen *U* and *W* satisfying condition (5.10). Using these values, a steady state value of state vector is obtained using continuous application of equation (5.9). Now vector V can be obtained using condition 2, which can be solved using simple linear equation system. After we have all three matrices, value of output y is calculated. After calculating output y, matrices *U* and *W* are modified using gradient descent, while satisfying condition (2). This particular procedure is one iteration. We repeat these iterations until output reaches equilibrium which is equal to y_0.

5.3 Fixed Point Constraint During Learning

Fixed point location and dynamics are the key invariant of nonlinear system, which should be included in the predicting RNN system. In this section the fixed point conditions derived in previous section are enforced during each step of continuous learning of RNN along the trajectory. This problem is approached as a constrained optimization point of view where the fixed point dynamics can be assigned as constraints while minimizing the mapping error over the whole time series. The problem is formulated as:

Minimize for U,V,W: $J(U,V,W) = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - Y_{Ri}\right)^2$ (5.19)

Such that :

$$V \cdot (U \cdot x_0 u + W \cdot z) = x_0 \tag{5.20}$$

$$\begin{bmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \\ \frac{\partial y_R}{\partial x}_{1x=x_0} & . & . & . & \frac{\partial y_R}{\partial x_n}_{x=x_0} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \\ \lambda_1 & . & . & . & \lambda_n \end{bmatrix} \tag{5.21}$$

$$\sum_i \sum_j w_{ij}^2 < 1/\max\left(f'\right)^2 \tag{5.22}$$

Where (5.20) and (5.21) are nonlinear equality constraints and (5.22) is an inequality constraint. In this work we enforce (5.20) and (5.21) using projected gradient descent

method. In this method the direction of descent also provides a constraint satisfying feasible point at each step, or the direction of descent is projected along constraint manifold.

In order to solve above equation, the let us first consider the general linear equality constrained problem

Minimize $f(w)$ (5.23)

Such that $Aw - b = 0$

Here A is a $r \times p$ matrix and b is r vector.

The gradient projection method is based on following argument [113]. Lets assume that w' is feasible, i.e. A direction $s, (\|s\| = 1)$ is sought such that a step αs from in the direction w' also satisfies constraint, i.e. $A(w' + \alpha s) - b = 0$. This implies

$$As = 0 \tag{5.24}$$

Moreover since $\|s\| = 1$ its follows that

$$1 - s^T s = 0 \tag{5.25}$$

Now it is required to calculate the steepest direction s such that it satisfies constraints (5.24) and (5.25). This is equivalent to determine s such that directional derivative at w':

$$\left.\frac{dF(\alpha)}{d\alpha}\right|_s = \nabla^T f(w')s \tag{5.26}$$

is minimized w.r.t. s, where $F(\alpha) = f(w' + \alpha s)$. By Lagrange theory for minimizing a function subject to equality constraints, we formulate the problem as lagrangian function

$$L(s, \lambda, \lambda_0) = \nabla^T f(w') + \lambda^T As + \lambda_0 (1 - s^T s) \tag{5.27}$$

where $s = [s_1 \quad , \quad , \quad , \quad s_p]$ correspond to cosine of s. The lagrangian conditions for constrained minimum are:

$$\nabla_s L = \nabla f(w') s + A^T \lambda - 2\lambda_0 s = 0 \tag{5.28}$$

$$\nabla_\lambda L = As = 0 \tag{5.29}$$

$$\nabla_{\lambda_0} L = (1 - s^T s) = 0 \tag{5.30}$$

Equation (5.28) yields,

$$s = \frac{1}{2\lambda_0} \left(\nabla f(w') + A^T \lambda \right) \tag{5.31}$$

Now by (5.29) and (5.30)

$$\lambda_0 = \pm \frac{1}{2} \left\| \nabla f(w') + A^T \lambda \right\| \tag{5.32}$$

Now substituting (5.32) in (5.31), we get $s = \pm \dfrac{\left(\nabla f(w') + A^T \lambda \right)}{\left\| \nabla f(w') + A^T \lambda \right\|}$

where we choose negative sign for maximum steepest descent. Thus the constrained direction of steepest descent is obtained as

$$s = -\frac{\left(\nabla f(w') + A^T \lambda\right)}{\left\|\nabla f(w') + A^T \lambda\right\|} \tag{5.33}$$

Now from (5.29) and (5.31), we get $A\left(\nabla f(w') + A^T \lambda\right) = 0$. Thus if $s \neq 0$ then $AA^T \lambda = -A\nabla f(w')$, which implies

$$\lambda = -\left(AA^T\right)^{-1} A\nabla f(w') \tag{5.34}$$

Substituting in (5.32) we get the final direction of constrained gradient descent as

$$s = \frac{-\left(I - \left(AA^T\right)^{-1} A\right)\nabla f(w')}{\left\|\nabla f(w') + A^T \lambda\right\|} \tag{5.35}$$

5.3.1 Extension to nonlinear constraint

The original RNN learning is an optimization problem with nonlinear constraints w.r.t weight parameters as shown in eq (5.19) with equality constraints (5.20), (5.21) which can be formulated in general form as

Minimize $f(w')$ (5.36)

Such that $h_i(w') = 0, \qquad i = 1,2,\ldots, r$

here $f(w') = J(U,V,W) = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - Y_{Ri}\right)^2$

and,

$$h_1(w') = V \cdot (U \cdot x_0 u + W \cdot z) - x_0$$

$$h_{i+1}(w') = \frac{\partial y_R}{\partial x}_{ix=x_0} - \lambda_i \quad \text{for } i = 1,2,\ldots,d$$

where the constraints are nonlinear. For solving this, the constraints can be linearized at every feasible point w' by Taylors expansion:

$$h_i(w) = h_i(w' + (w - w')) \cong h_i(w') + \nabla^T h_i(w')(w - w') \tag{5.37}$$

which allows for the following constraint:

$$\nabla^T h_i(w')(w - w') = 0, \qquad i = 1,2,\ldots,r \tag{5.38}$$

in the neighbourhood of w', since $h_i(w') = 0$. This is written in matrix form as

$$\left[\frac{\partial h(w')}{\partial w}\right]^T w - b = 0 \tag{5.39}$$

which is equivalent to the linear constraint problem (5.20)

where $A = \left[\frac{\partial h(w')}{\partial w}\right]^T$ and $b = \left[\frac{\partial h(w')}{\partial w}\right]^T w'$. (5.40)

(5.37) implies that matrix A is dependent on feasible point w'. However due to nonlinearity, the approximation of linearization is applicable only in a small neighbourhood of w', which can result in condition $h_i(\overline{w_1}) \neq 0$ where $\overline{w_1} = (w' + \alpha s)$ causing failure in satisfaction of constraint. A correction step $\overline{w_1} \rightarrow w_1$ is incorporated, such that the projection at $\overline{w_1}$ and w_1 is same, or

$$\left(I - A^T\left(AA^T\right)^{-1}A\right)\left(w_1 - \overline{w_1}\right) = 0 \tag{5.41}$$

Now since $h(w) = Aw - b$ for both $h(w_1)$ and $h(\overline{w_1})$, and given the fact that $h(w_1) = 0$, we get

$$w_1 = \overline{w_1} - A^T\left(AA^T\right)^{-1}h\left(\overline{w_1}\right) \tag{5.42}$$

which is the actual next step in gradient descent.

5.4 Local Jacobian Learning

Learning of Jacobian is beneficial in neural network training for learning of transient information hidden in time series. Moreover, learning of gradient of time series along with input output vector mapping enable better generalization. The merits of learning of jacobian along with the actual input output mapping has been demonstrated with examples in few works such as [114]. In this work we test if simultaneous jacobian learning assists in improving results in case of electricity price time series forecasting. In this section we employ penalty based formulation of gradient descent method for simultaneous learning of

mapping and its jacobian. This error correction learning rule and the corresponding weight adjustment can be formulated as

$$\Delta w = \Delta w_m + \Delta w_j \tag{5.43}$$

Where Δw_m is the weight adjustment for minimizing mapping error derived in the previous section using projected gradient descent, and Δw_j is the weight correction corresponding to minimization of jacobian error. This implies,

$$\Delta w_m = -\eta_m \frac{\delta E_m}{\delta w} \text{ and } \Delta w_J = -\eta_J \frac{\delta E_J}{\delta w} \tag{5.44}$$

where E_m is the mapping error and E_J is the jacobian error, while η_m and η_J are the corresponding learning rates.

We can derive the second term in (5.44)

$$E_J = \frac{1}{2}\sum_n \left(v_n - \hat{v}_n\right)^2 \text{, where } v_n = \frac{\partial y}{\partial x_n} = \sum w_k f_k^{'} u_{nk} \text{ ,} \tag{5.45}$$

and $\hat{v}_n$ is the actual gradient of time series with respect to state *n* which is calculated for each point in the trajectory using the method employed for calculating FTLE in section 4.5.

5.5 Summary

In this chapter a novel learning approach is proposed for training recurrent neural networks in complex scenario such as electricity price time series. The invariant dynamics of time series are incorporated during the learning process (in particular fixed point dynamics) and

enforced during each step of learning using gradient projection method. This approach was proposed in attempt to achieve a dynamic RNN model which learns the inherent complex nonlinearities of the system which makes it a closer approximation of actual nonlinear system. This approach can also be seen as using heuristic information to achieve global optimum solution during time series based RNN learning. The three steps of learning process include intelligent weight initialization using fixed point dynamics, enforcing fixed point conditions during each learning step and enforcing local jacobian during the course of time series. This approach is expected to perform with high accuracy in case of complex but deterministic systems. The developed RNN model is named as PGRNN (Projected Gradient based RNN). The implementation results of PGRNN are shown in chapter 8.

While on one hand this approach brings the trained model closer to exact nonlinear dynamics of time series, the chances of unsuccessful learning are high in case of slight nonstationarity. Moreover the discussion in chapter 4 hints towards occurrence of nonstationarity at slow scale. Considering this scenario, multiple scale dynamics based approach has been adopted in neural networks modelling which has been discussed in chapter 6 and 7.

Chapter 6

Electricity Price Time Series Prediction Using Hybrid RNN-FHN model

In this chapter firstly the presence of multiple scale dynamics in the electricity price time series system is affirmed. The approach adopts extraction of dynamics of slow parameter system, variations in which affect the behaviour of fast actual time series [115]. Further a dynamic system with slow and fast scales, namely Fitz-Hugh Nagumo (FHN), is used and hybridized with recurrent neural networks. The property of the multiple scale equation system allows the mechanism of spiking in such regimes. In turn, the parameters and coupling variables of this excitable system are determined using an RNN based model. As a result the learned hybrid model would achieve a desired level of modelling accuracy. Evolutionary Strategies (ES), an evolutionary computation technique, has been emplyed for training the feed-forward and feedback weights of the RNN network, and the trained overall model is applied for forecasting in deregulated markets. The developed hybrid model was tested in various markets worldwide over different seasons to test its forecasting ability, adaptability and robustness. The developed approach is tested over the most volatile deregulated markets in hour-ahead and day ahead scenario. Extensive comparative studies suggest that our approach yields favourite results in both point and interval forecasting.

It should be noted that in this chapter hour ahead spot and interval forecasting has been presented in this paper. Although day-ahead forecasting is more important

This chapter is organized as follows. In the next section we describe and extract the multiple scale characteristics of electricity price time series. Section 6.3 briefly describes behavior of FHN is slow and fast time scales and presents the hybrid model. Section 6.4 presents implementation over various markets worldwide and the results corresponding to

hour ahead forecasting (results for day ahead forecasting are given in chapter 7). Section 6.5 presents discussion followed by conclusion in section 6.6.

6.1 Multiple Scale Dynamics in Electricity Price Time Series

Most of the current real world physical systems share a basic mechanism, where a hidden slowly evolving process leads to variation in the systems process evolving in a faster time scale. The slow dynamics may correspond to gradually evolving system parameters. Or at micro level, in competitive scenarios, multiple scales can “enter” the system when financial agents interact at different scales. Prices are endogenously formed in a financial market by the aggregate trading decisions of a large number of investors who interact by trading assets, while the resource money is limited. While dealing with huge amount of money for the same asset, these agents differ with respect to the time scale in which take trading decisions. These different time scales significantly impact the dynamics of price. When the dynamics of market prices is modeled on a macro-level as the result of the dynamic coupling of two dynamical components, the degree of their dynamical decoupling is shown to have a significant impact on volatility clustering, and the multifractal behavior of time series.

The primary motivation in this work is to study the evolution of parameter changes in the slow scale and prediction of changes. In this section we extract the slow and fast scale characteristics of electricity price time series, which in turn will help in justifying FHN based modeling of time series. In FHN model, the spike formation is possible because of presence of these two time scales, slow evolution along an invariant manifold and instability of fast dynamics transverse to this manifold. The location of these spikes is fixed and can be determined from an analysis of the dynamics on the slow manifold. We are interested in developing methods capable of tracking the evolution of slow damage states in real time.

This will provide the hidden time series dataset that will shed light on poorly understood multi-scale dynamics of process. The approach tracks the evolution of slowly evolving hidden state using only measurement of fast subsystem which is source of non-stationarity on long time scales [116].

$$\begin{aligned} \dot{x} &= f(x, l(\omega), t), \\ \dot{\omega} &= \varepsilon g(x, \omega) \end{aligned} \tag{6.1}$$

where x is the fast observable variable, is the slow dynamic variable which is not directly accessible and behaves as the parameter whose variation causes nonstationarity in fast variable. is the function of parameter vector , t is the time, the constant is the separation of time-scale between fast and slow dynamics, and y is the scalar quantity derived from fast variable x via function h.

We implement the algorithm proposed by Chelidze et. al. [116] based on extraction of slow variable features based on fast variable dynamics. The method uses idea of phase space warping which refers to distortions in fast dynamics phase space and uses short-time reference modeling error as its primary measurement.

Basically reconstruction of slow time trajectory based on fast time measurement is done based on a metric which is sensitive to change in the dynamics caused by variation in slow parameter. This metric can be viewed as measure of deformation of fast time phase space due to slow time drifts. This metric used by Chelidze et. al [116] is called Phase Space Warping Function (PSWF).

Scalar data of fast measurements is collected and phase space is reconstructed using delay coordinate embedding as

$$\overline{y_n} = [y(n), y(n+1), ..., y(n+d-1)]^T \in R^d \tag{6.2}$$

Here d is the embedding dimension. PSWF is formulated as

$$e_R(\phi; \overline{y_n}) = \overline{y_{n+1}} - P(\overline{y_n}; \phi_{damage}) \tag{6.3}$$

Here y_n is a point in the reconstructed fast phase space. $P(\overline{y_n}; \phi_{damage})$ is the prediction of fast state for next time step based on a linear prediction model P.

The procedure for evaluating PSWF follows the partition of phase space into multiple disjoint hypercuboids, C_i, *i=1,…,N* which are small enough such that linear approximation is applicable. The PSWF value is calculated in each such region as

$$e_i(\phi) = \frac{1}{N_{ci}} \sum_{y \in C_i} \hat{e} \tag{6.4}$$

Here *Nci* is the number of points in hypercuboid C_i, using data based estimated linear model P. All of these averaged PSWFs are collected as *dN*-dimensional feature vector $[e_1; e_2; ...; e_N]$ for each stack of observation, or *{j} j=1,..,q* snapshot of observation. In order to examine if the phase space of slow parameter can be reconstructed without a priori knowledge of the changes occurring in damage states, The authors stacked the feature vectors as a time sequence in a row vector of a tracking matrix $T \in R^{q \times dN}$. Then the smooth orthogonal decomposition (SOD) of this tracking matrix provides the reconstructed damage phase space (damaged by slow variable variation) which is an approximation of actual damage phase space but is in affine relationship with it. Here the SOD is performed via generalized singular value decomposition (GSVD) of tracking matrix and its derivative DT [116].

$$T = UCX^T \quad ; \quad DT = VSX^T \tag{6.5}$$

Upon GSVD [116], the smooth orthogonal coordinates are given by the columns of *UC* while smooth orthogonal modes of damage phase space are provided by columns of X^{-1}. Smooth orthogonal values SOVs are obtained as $\sigma = diag(C^T C) / diag(S^T S)$.

In this work we implemented the above procedure on electricity price time series to obtain the damage caused by hidden parameters in the system. These act as the variables which cause sudden transitions in time series such as spiking. The time series from California and Ontario market was taken as reference to study the damage evolution. Each time series was observed via moving window of snapshots containing 300 patterns each. Each observation was divided into 35 neighborhood hypercuboids. The corresponding SOVs for Ontario market are shown in fig. 6.1 below.

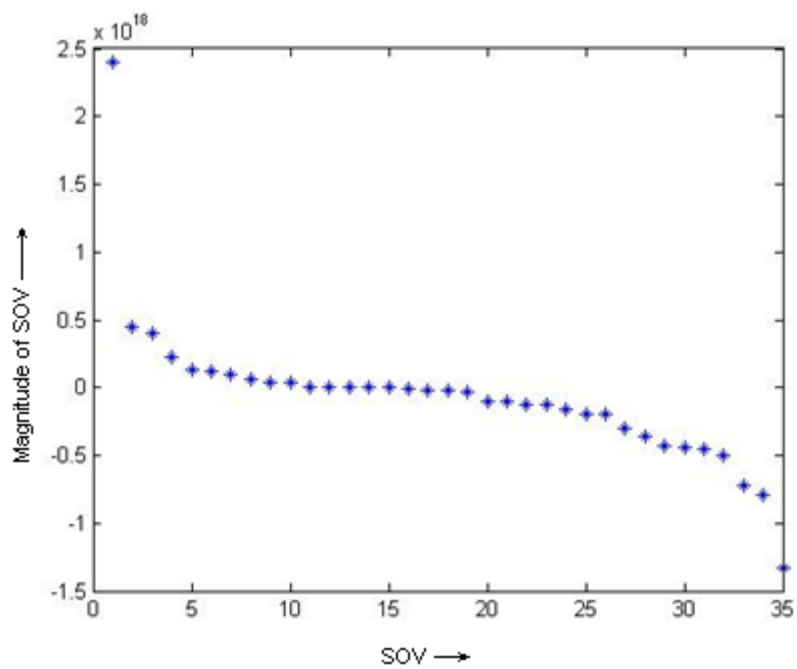


Fig.6.1 Smooth Orthogonal Values of slow subsystem

It can be observed that the damage phase space dynamics are dominated by single variable. The SOD analysis shows that the there is only one SOV that is substantially larger than the rest. More importantly, the presence of additional time scale in this system is verified.

Furthermore, the SOMs corresponding to this SOV is plotted along with the actual time series as shown in Fig. 6.2 below. Interestingly it can be noted that the variation in parameter value take place a little ahead of actual tipping state in time series. This could greatly facilitate in prediction of critical transition in these time series systems.

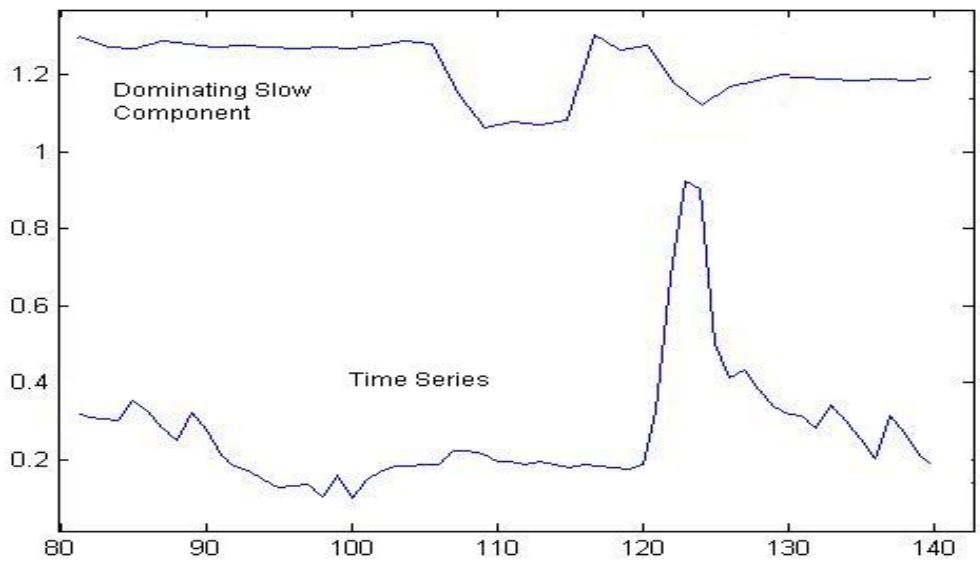


Fig. 6.2 Dominating Slow Component vs Time Series

The presence of multiple scales in this system encourages us to use a two-scale system for modelling this time series. Famous two scale Fitz-Hugh Nagumo (FHN) system has been employed for this task. This is further hybridized with recurrent neural networks, given their universal nonlinear approximation capability, to successfully model this system. The hybrid model is described in next sections.

6.2 Fitz-Hugh Nagumo Model

Dynamic models are extremely difficult to model, especially when incorporating intermittent spikes of variable height along with stable region dynamics. While, in general jump processes are not ideal choice for modeling market prices as they face the problem of mean reversion and inaccurate modeling capability in stable region. In this work, both approaches have been hybridized for accurate prediction on overall time series. FHN system has been taken as the basic mathematical model for spiking. Under suitable choice of parameters, FHN operates in resonating regime which works well for power markets [117]. FHN model represents a wide class of non-linear excitable oscillatory systems [117-118] and is described using following equation.

$$\begin{aligned} \dot{u} &= f(u,v) = (ku - \lambda u^3/3 - v)/\varepsilon + I, \\ \dot{v} &= g(u,v) = (\gamma u + b - \beta v - A\sin(\omega t)) \end{aligned} \tag{6.6}$$

Here $\varepsilon, \gamma, \lambda, \beta, A, \varpi, k, b$ and I are equation parameters. Variations in these parameter values lead to “Bifurcation” phenomenon of behaviour transition in dynamics of the system. A bifurcation occurs when a small smooth change made to the parameter values (the bifurcation parameters) of a system causes a sudden 'qualitative' or topological change in system behaviour. Detailed study of this phenomenon can be found in [118]. In (6.6), u behaves as an excitable variable and v is the slow refractory variable and I is an external variable. FHN model follows a symmetric property when $I = 0$, namely, (6.6) is invariant under transformation with respect to u and v. With $I \neq 0$, the symmetric property is broken. The system exhibits topologically different dynamics for different values of bifurcation parameters. There are at least two time scales present in eq (6.6) due to spike refractory period (which corresponds to internal frequency of the system) and the sinusoidal

forcing frequency. The interplay between these time scales result leads to multiple scale behaviour which is a typical source of intermittent spikes in a dynamical system. Fig. 6.3 depicts the behaviour of eq(6.6) for certain parameter values. It can be observed that the dynamics match typical power market.

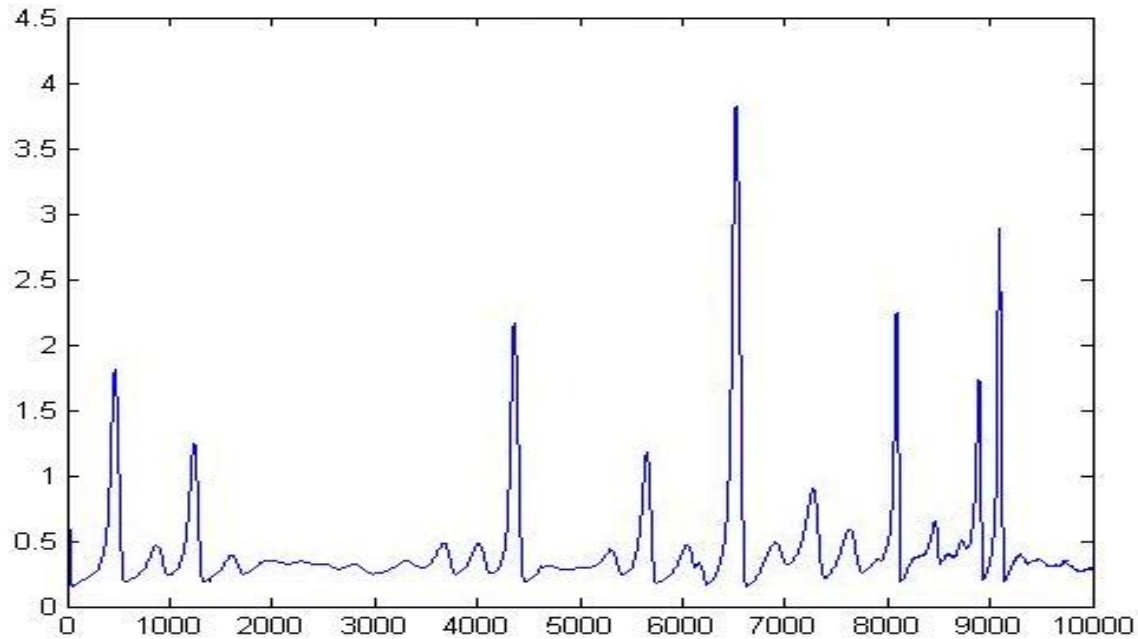


Fig.6.3. Time response of FHN for ε=0.11, k=1, λ=1, b=2, γ=1, β=1, A=2, ωf=2 and *I*=1.

6.3 Proposed Model

The proposed approach consisted of three components; a FHN model to mimic spiking behaviour, a RNN unit to regulate FHN, and feedforward neural network to model the residue of RNN-FHN.

Hybrid RNN-FHN Model

Desired behaviour is achievable upon intelligent control of the bifurcation parameters $\varepsilon, \gamma, k, \lambda, b, \gamma, \beta, A$ and ω are modulated using RNN according to the input values from the past time series which reflect market conditions. The motivation behind using RNN for

this task is its dynamic nature and nonlinear approximation capabilities [23]. RNN offers explicit modeling of time and memory and exhibits temporal behaviour which makes it pertinent for this dynamic problem. In this work, Elman state feedback configuration of RNN has been used, as shown in Fig. 6.4.

Equation (6.7) presents the discrete version of the FHN model obtained by Euler discretization.

$$\begin{aligned} u(k+1) &= u(k) + \partial(ku(k) - \lambda u^3(k)/3 - v)/\varepsilon, \\ v(k+1) &= v(k) + \partial(\gamma u + b - \beta v - A\sin(\omega t)) \end{aligned} \tag{6.7}$$

Here ∂ is the discretization step size. The value of ∂ was chosen to be 0.05 in this work for good approximation of the continuous model and providing stability to the iterations, and I was chosen as 1.

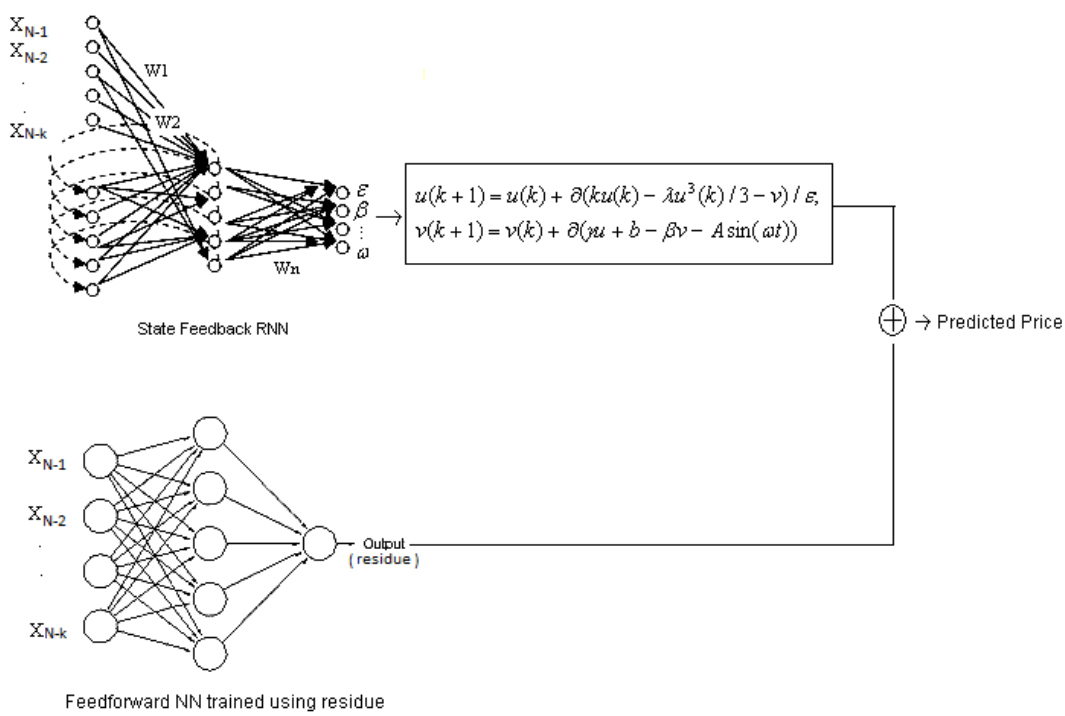


Fig.6.4 Proposed Hybrid Model with RNN and FHN coupled system

Fig. 6.4 illustrates the proposed strategy. RNN uses K previous time series values X_{N-1}, X_{N-2} …, X_{N-K} as inputs. W_i , $1 \le i \le n$ are the complete set of weights connecting input layer to

hidden layer, hidden layer to output layer and feedback weights connecting hidden layer to input layer. One training vector is, therefore, of length *n*. RNN evaluates the bifurcation parameter values. Subsequently, FHN system of equations uses these calculated bifurcation parameters and obtains the future value in time series X_N. The learning in this model is carried out through training of RNN weights, using evolutionary strategies, for the network to approximate the relationship between time series values and bifurcation parameters, as explained in the next subsection.

Feedforward neural network used to predict the residual error of uses *K* previous time series values as inputs. The output of this network is added to the predicted output of RNN-FHN. The network is trained using backpropagation algorithm. The overall hybrid model and complete strategy are elaborated in Fig.6.4.

6.4 Training of RNN in Hybrid Model

Evolutionary Strategies ([119]) provide a promising algorithm for parametric learning of RNN. These algorithms are distinguished by their reliance on a population of search space positions, rather than a single position to locate the extrema of a function defined over a search space. Using random search operators such as mutation and recombination, and probabilistic selection, these algorithms implement a non-monotonic search that performs better in complex multimodal environments, which is pertinent to this forecasting problem. Derivation of a learning procedure in this hybrid problem can be a rigorous task for numerical and gradient based techniques, while in contrast ES is representation independent. The main motivation behind using ES is that it is a global optimization method which scales well to high dimensional problems, while most of other learning algorithms run a risk of premature convergence to local extremum in complex problems.

The objective is to evaluate all connecting weight values of the network shown in Fig.

6.4. Each chromosome, which represents the solution weights, can be represented as an array of weight values along with the mutation parameters for the individual weights. The weight vector chromosome representation is shown in Fig. 6.5.

W1	W2	.	.	.	Wn	$\sigma 1$	σ_2	.	.	.	σ_n

Fig.6.5. Chromosome Representation

Here mutation parameter σk corresponds to weight W_k. Mutation of weights takes place as given in (6.8) below.

$$\sigma(k+1) = \sigma(k) * \exp(\tau * N(0,1) + \tau' * N'(0,1))$$
$$W(k+1) = W(k) + \sigma(k+1) * N(0,1) \tag{6.8}$$

The mutation scheme used in this work is adaptive, uncorrelated mutation scheme as shown in (6.9). Here N(0,1) is a Gaussian variable with 0 mean and variance equal to 1. τ and τ' are learning rate parameters which are determined as

$$\tau = 1/(2m)^{1/2}$$
$$\tau' = 1/(2m^{1/2})^{1/2} \tag{6.9}$$

where m is the length of the chromosome. Here the mutation works on coordinate level which allows the algorithm to adapt the nature and scale of its search to learning task. More details on mutation mechanism can be found in [119]. During one search cycle, members of the population are ranked according to a fitness function, and those with higher fitness are probabilistically selected to become parents. A population of λ children is reproduced from a population of μ parents. Selective numbers of these parents or children constitute the population for the next generation which takes part in the next iteration in training procedure. This selection process is crucial for the effectiveness of algorithm.

In the proposed strategy, non-elitist, or (μ , λ) selection scheme was used, allowing good exploration of search space and averting convergence to suboptimal solutions [119]. The fitness of each child is inversely proportional to the error evaluated over complete training data. Thus lowering the training error (obtained using that child chromosome) improves the fitness of the child and the chances of it being selected for the next generation.

Table 6.1
Evolutionary Strategy Parameters

Mutation Type	Adaptive
Crossover Type	No Crossover
Population Size(μ)	10-20
Children Size(λ)	8*Population size
Selection	Non-Elitist (μ, λ) strategy

The parameters of ES applied in this work are shown in Table I. Population size and the numbers of offspring were chosen to be high to facilitate good exploration of solution space. The overall strategy is illustrated in Fig. 6.4. Various hyper-parameters such as the size of recurrent neural network, RNN topology (number of inputs and size of middle layer), population size in evolutionary strategy, children size, selection scheme etc. were chosen based on extensive experimentation

6.5 Prediction of Hourly Prices

The proposed hybrid model was implemented for forecasting of hourly prices in testing data. Time series data from each market was divided into training and testing sets. Training of hybrid model was carried out in supervised manner and the performance of trained model was examined on unseen test data, without any adaptation.

6.6 Training and Testing Data

The electricity markets most affected by restructuring include California and PJM markets in the US, UK, Spain, Ontario market in Canada, and NEMMCO market in Australia. In some of these markets, higher energy prices have enticed new generating companies to enter, resulting in a greater competition and higher volatility [120], while in markets like Ontario, huge networks with neighbor markets and regulated utilities significantly complicates the competition scenario. Volatility analysis done in some of the previous works has concluded that UK and Ontario markets exhibit highest degree of volatility compared to other deregulated markets worldwide [120]. We have presented the implementation results in these five markets. The trained model was implemented on unseen testing datasets given in Table II. The training dataset includes one month data prior to testing period. The datasets used incorporate the defining characteristics of electricity price time series, like high volatility, occasional extreme jumps, and volatility clustering.

Table 6.2.
Testing Dataset Used

Market	Season	Testing Data
Ontraio(week 1-2)	Spring	Apr 26 – May 9, 2004
Ontario(week 3-4)	Summer	Jul 26 – Aug 8, 2004
Ontario(week 5-6)	Winter	Dec 13 – Dec 26, 2004
NEMMCO Victoria	Summer	Sep 01 – Sep 07, 2004

NEMMCO NSW	Fall	May 21 – May 27,2006
Spain	Summer	Aug 19 – Aug 25, 2002
California	Winter	Dec 26 – Jan 9, 1999

In normal circumstances, range of dataset is normalized from 0.1-0.9 with 0.9 corresponding to maximum value in data and 0.1 corresponding to minimum. However presence of large and rare spikes can affect the data characteristics upon normalization. Thus we incorporate a special normalization procedure where high value spikes are ignored while normalizing the range of dataset, as shown in Fig. 6.6.

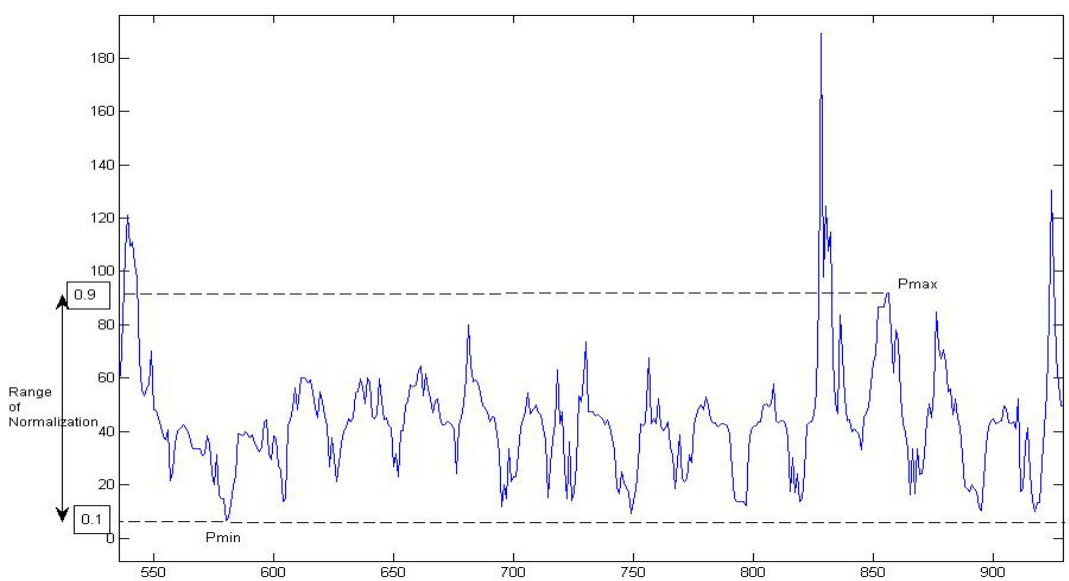


Fig.6.6. Selection of range of normalization for spiky time series. The high value spikes are neglected.

$$d = \left(\frac{P_{\max} - P_{\min}}{8} \right); \tag{6.10}$$

$$P_i = \frac{\left[P_i - \left(P_{\min} - d \right) \right]}{\left[\left(P_{\max} + d \right) - \left(P_{\min} - d \right) \right]}$$

The normalization is carried out as in (5), where $P_{\max}$ and $P_{\min}$ are selected as shown in Fig.

5. In this work, $P_{\max}$ is chosen after neglecting 10% of the highest prices to avoid spikes. Output of (5) is the desired normal price. It should be noted that normalized prices in spiking region exceed normalized $P_{\max}$. Application of the proposed model on these training and testing data is presented in the next section.

6.7 Experimental Results

MAPE (Mean Absolute Percentage Error) has been chosen as the error metric due to its wide application in evaluating forecasting models. MAPE is defined as:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|P_i - \overline{P}_i\right|}{\overline{P}_i} \tag{6.11}$$

Here *N* is the total number of data, $\overline{P}_i$ is the actual price and P_i is the predicted price. Many texts use variants of MAPE, namely mean absolute daily error (MADE), mean absolute weekly error (MAWE) etc. as forecasting error indices as low prices can have adverse shooting effects on MAPE. However for fair comparison with other works on hourly prices we use MAPE in this work as a performance index for hour ahead forecasting.

Section 6.6 presents the performance of the hybrid model over point forecasting of hourly prices compared to other approaches recently proposed in the past, and relevant benchmarking is performed. Application of evolutionary algorithms as the training procedure brings large computation time to this developed model. However this computation time is not a key matter of concern as the learning process is carried out in offline manner and the trained model can be applied on a large set of unseen test data.

The corresponding performance curves, which show the comparison of predicted and actual results, are shown in Fig.6.7 for Spanish and Ontario market.

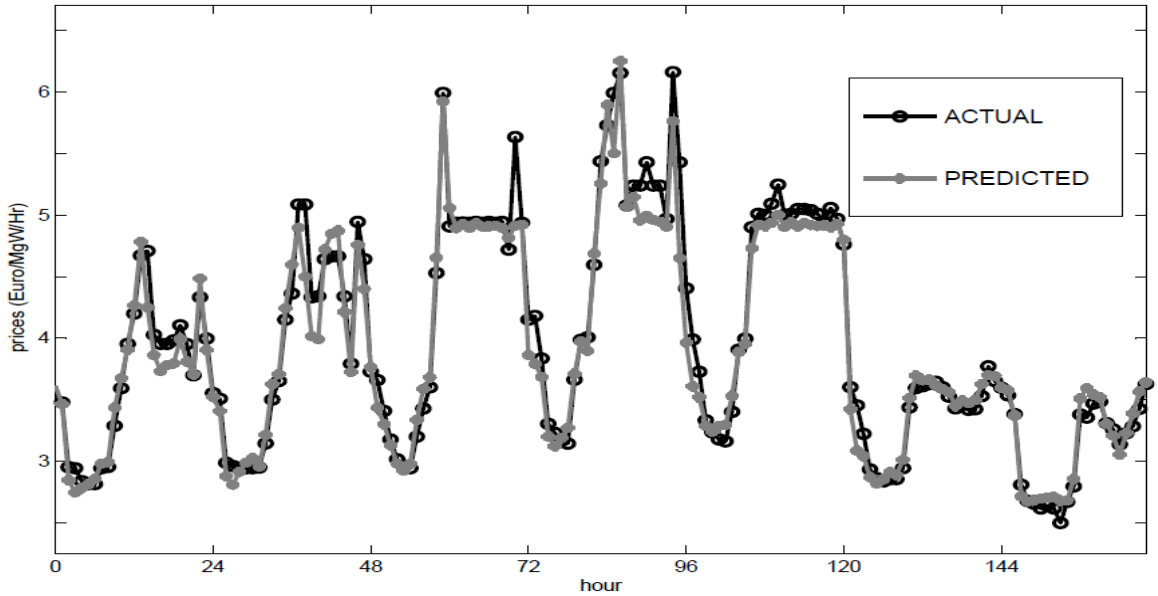


Fig. 6.7(a). Actual and Predicted curve in Spanish market during Summer

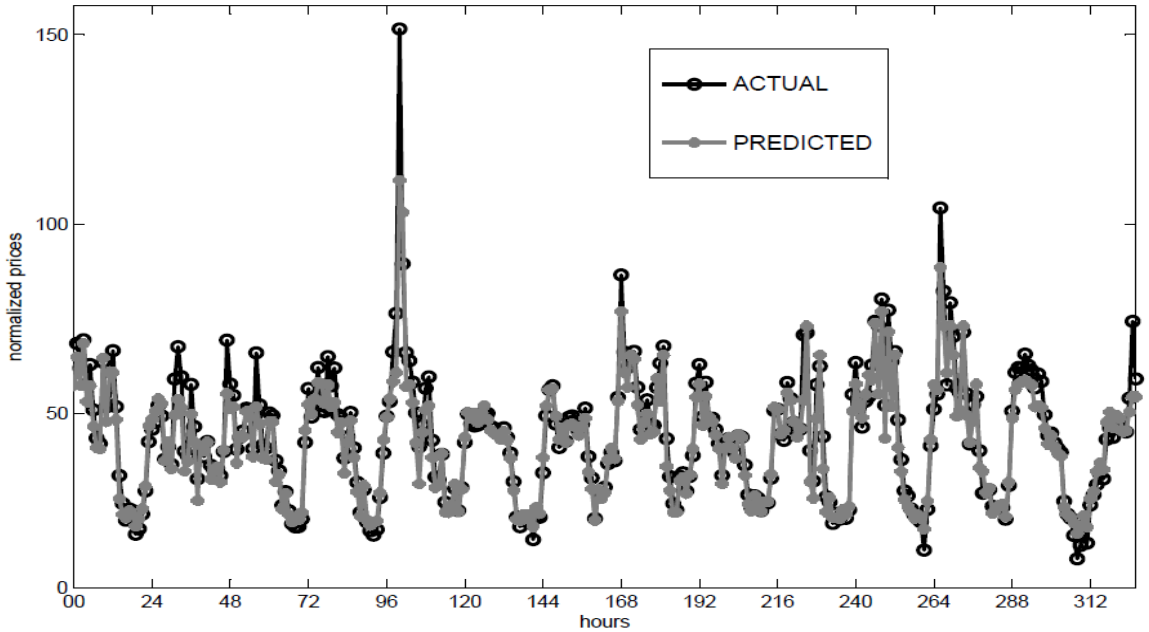


Fig. 6.7(b). Actual and Predicted curve in Ontario market during Spring

These performance curves depict the ability of the proposed model of providing close approximations over the stable regimes of time series while satisfactorily tracking the spiky regions. It can be seen that market clearing price (MCP) in Spanish market are stable and proposed approach closely approximates these prices, except 3rd (Tue) and 4th (Wed) day

of the week. While hour-ahead energy prices (HOEP) in Ontario market are comparatively more volatile with presence of intermittent spikes. Fig. 6.7(b) indicates that proposed approach performs satisfactorily in these spiking regimes (except day 5) and also in stable regimes (except day 2-3). Prediction results in terms of MAPE are presented in Table III, where comparison with benchmark models is given.

The studies carried out in the past for prediction in Spanish, California, Ontario and Australia markets are given in [19, 120-126]. In [120], [121] and [124], one hour ahead HOEP forecasts have been reported using multi-adaptive regressive spline (MARS), RNN trained using expectation maximization procedure and nonlinear neuro-fuzzy model respectively. In [127], Spanish hourly prices have been modeled using hybrid NN-wavelet model; and using input-output hidden markov model in [123]. A hybrid approach based on NN-ARIMA and wavelet-TAR model has been applied for modeling hour-ahead prices in NSW market [126, 128]. The benchmark model used for comparison in Victoria and California market includes NN based model [19, 129]. The corresponding results are compared with the proposed approach in Table 6.3. Here, column 3 includes benchmark models used for direct comparison with proposed approach. For fair comparison, the testing periods used for proposed work are exactly same as those used in benchmark models given in column 3. While in column 4, other approaches applied in past over similar datasets (but different test periods) in same markets are given.

Table 6.3
Performance Comparison over Deregulated Markets

Market	Proposed Model	Benchmark models	Other Approaches
Ontario Season1 (week1,2)	10.76%	15.12% [122]	12.25% [120]/ 19.83% [124]
Ontario Season2 (week3,4)	9.12%	10.31% [122]	10.20% [120]
Ontario Season3 (week5,6)	11.61%	15.74%[122]	12.85% [120]
Nemmco Vic	8.24%	10.69% [19]	-
Nemmco NSW	8.33%	15.00% [128]	10.00%[126]
Spain	8.87%	9.50% [125]	10.38% [122]/ 15.83 [123]
California	5.32%	7.932[129]	-

Performance curves in Fig. 6.7 and the comparison table 6.3 depict the forecasting ability of proposed model. It can be deduced that the proposed hybrid model is able to

approximate the dynamics of hourly electricity prices with satisfactory degree of accuracy. It is observed that the proposed approach outperforms the methods suggested in past on specified test set. Ability of the developed model to satisfactorily predict seemingly sporadic spikes allows the market participants to reduce financial losses in the high risk zone.

6.8 Interval Forecasting

In addition to forecasting the price value, predicting the distribution of the future price, i.e., the prediction interval (PI) is highly meaningful. Price intervals effectively reflect the uncertainties in the predication results and thus the risk associated with forecasting model. Prediction intervals are important in time series forecasting as they-

a) Allow to assess future uncertainty,
b) Enable different strategies to be planned for different outcomes indicated by the P.I.
c) It can be employed to evaluate the risks of the decisions made by market participants.

Interval forecasting includes proposition of a prediction interval $[l_i, u_i]$, shown in Fig. 7, or future values with a specific confidence level (α -level) in that interval [130-131].

Interval forecasts with confidence interval (CI) of 90%, 95% and 99% were obtained using the proposed approach for deregulated markets by following the statistical methodology [132] where the mean μ and variance σ of forecasts are obtained upon m multiple runs on testing data. The mean μ is the mean of m predictions and σ is the variance of those predictions. CI is obtained as

$$CI = \mu \pm z(\sigma / \sqrt{m}) \tag{6.12}$$

Here z = 2.576, 1.96 and 1.645 for α =99%, 95% and 90% respectively.

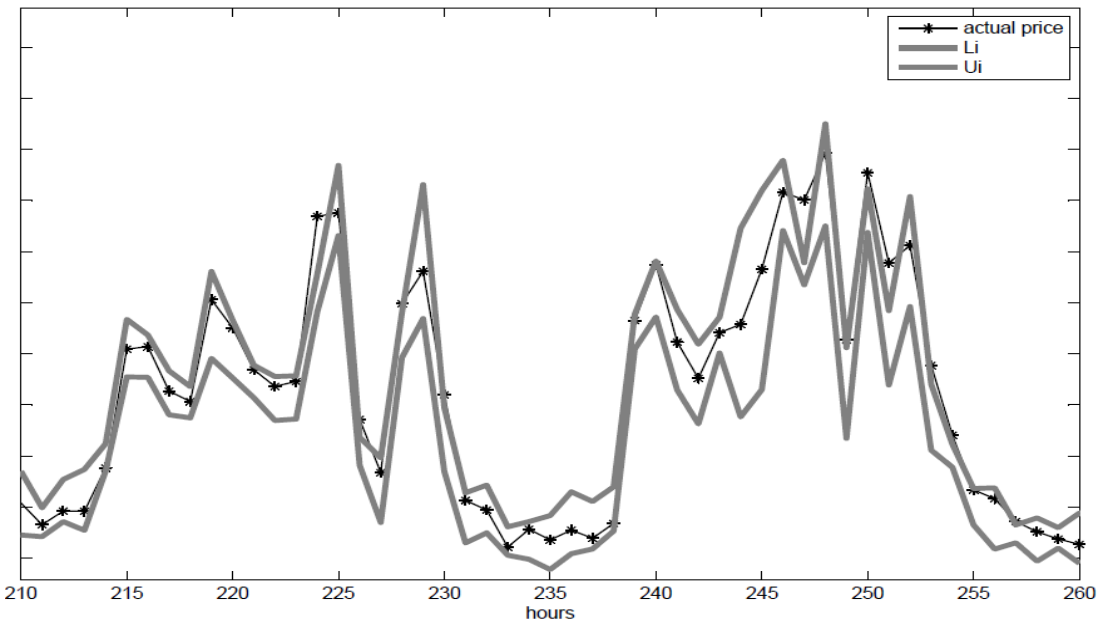


Fig.6.8(a). Enlarged section of Interval Forecasts for Ontario market

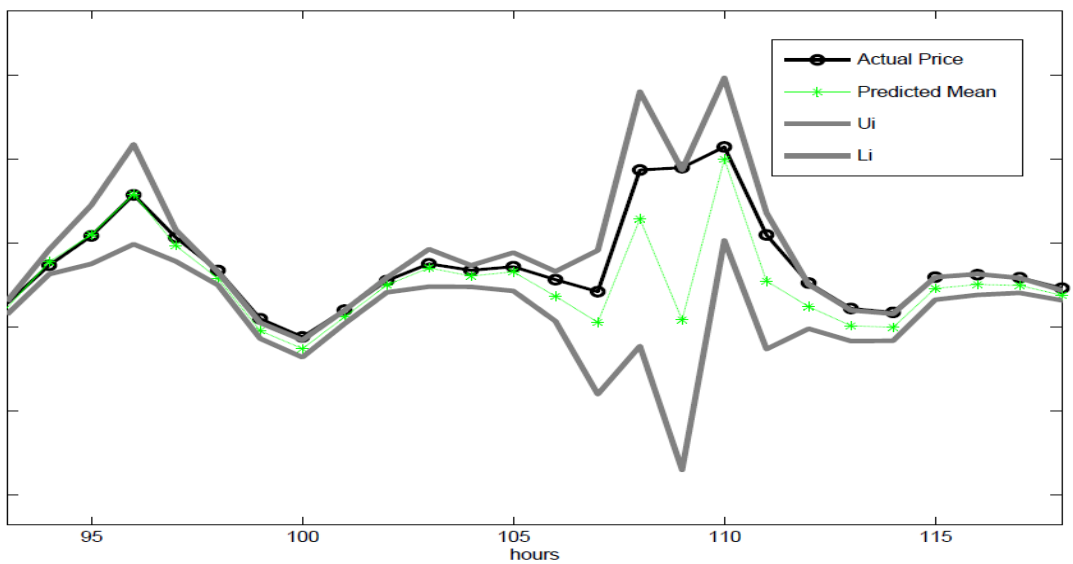


Fig.6.8(b). Enlarged section of Interval Forecasts for California market with large forecasting range around spiky region.

The criteria of absolute coverage error (ACE) [130], conditional coverage (CC) and unconditional coverage (UC) [133] have been used to evaluate the obtained intervals. Given N observations $y_i, 1 \le i \le N$ of time series and corresponding interval $[l_i, u_i]$ with CI= α

$ACE = \alpha - \hat{\alpha}$; where

$$\hat{\alpha} = \frac{freq(y_i \varepsilon [l_i, u_i])}{N} \tag{6.13}$$

Thus $\hat{\alpha}$ is the fraction of observations lying within the interval.

Fig. 6.8(a) shows the enlarged section of time series along with forecasted prediction intervals with CI=95% for Ontario market. It desirable to achieve intervals with smaller range $[l_i, u_i]$, while maintaining low value of ACE. Similar criteria of unconditional coverage, LRuc, indicates the quality of coverage [133], as shown in (6.14).

$$LR_{UC} = -2\log\left[\frac{p^{n1}(1-p)^{n0}}{\pi^{n1}(1-\pi)^{n0}}\right] \tag{6.14}$$

$$\pi = \frac{n1}{n1+n0}$$

Here p is the coverage rate; *n1* is the number of exceptions for the proposed prediction interval; n0 is the number of non-exceptions.

In addition of having a good absolute coverage, a good interval forecast should be narrow in stable periods and wide in volatile spiky regions, indicated by test of conditional coverage (CC) [133]. This property allows uniformity of coverage. To assess this property for interval forecast obtained from proposed model, Pearson-type χ^2 test has been used [134]. This test is asymptotically equivalent to likelihood ratio test suggested in [134]. The chi-square statistic is evaluated as [135].

$$LR_{CC} = -2\log\left[\frac{p^{n1}(1-p)^{n0}}{\pi_{01}^{\ n00}(1-\pi_{01})^{n01}\pi_{11}^{\ n10}(1-\pi_{11})^{n11}}\right] \tag{6.15}$$

here nij is the number value i followed by j in indicator sequence; and

$$\pi_{01} = \frac{n01}{n01 + n00}; \qquad \pi_{11} = \frac{n11}{n10 + n11} \tag{6.16}$$

The statistics ACE, LRuc and LRcc are given in Table 6.4. Low values of ACE indicate that the proposed CI has coverage and confidence level closed to desired value. This suggests that quality of one-hour-ahead PI obtained using proposed approach is good. ACE and unconditional coverage statistic LRuc for Ontario are higher compared to other markets due to higher volatility in this market. Moreover, it can be observed in Fig.6.8(b) that prediction interval are narrow in stable regime while interval range is large in spiking region which is due to high uncertainty associated with the regime. This property suggests uniformity of coverage over testing dataset. However it can be observed in Table IV that conditional coverage statistic vary significantly from LRuc implying that coverage uniformity is not as good as indicated by Fig. 6.8(b).

Table 6.4.
Interval Forecast

Market	ACE			LRUC			LRCC		
	CI =90	CI =95	CI =99	CI =90	CI =95	CI =99	CI =90	CI =95	CI =99
Ontario Season1 (week1,2)	6.2	5.33	5.3	33.4	28.4	28.27	36.2	32.54	31.6
Ontario Season2 (week3,4)	7.4	8.4	7.2	33.3	38.2	32.0	34.2	40.46	34.3
Ontario Season3 (week5,6)	4.5	6.2	5.7	20.2	28.4	25.8	22.6	30.56	28.37
Nemmco Vic	5.1	5.6	6.4	24.2	24.8	26.86	27.45	28.7	29.12
Nemmco NSW	3.3	5.28	4.2	15.7	22.0	28.2	18.1	23.0	31.7
Spain	5.2	4.7	4.5	19.3	22.3	26.5	21.7	24.1	27.4
California	2.2	2.3	2.5	16.8	18.5	19.7	18.8	21.73	20.57

Statistical Analysis

Fig.6.9 depicts the distribution of percentage errors corresponding to the predicted forecasts. Each bar of the histogram corresponds to the number of forecasts in that error range. It can be seen that height of bars is maximum around minimum error, while there are

very few bars with low height in high error value range. This scenario is favourable in reducing eminent financial risks associated with high forecasting error which is important in spiky time series scenario.

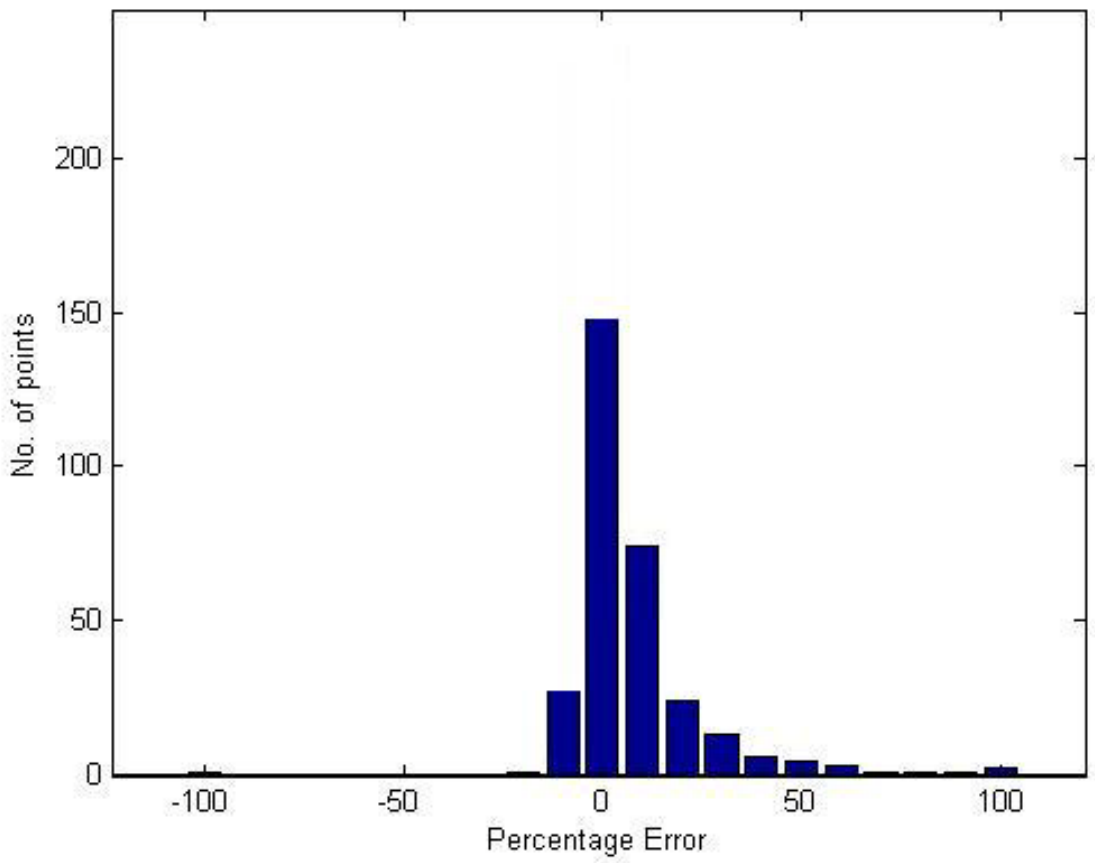


Fig.6.9. Histogram for error values of California market

Table 6.5.
Error Statistics

Market	Min(%)	Max(%)	Mean(%)	Std. Dev(%)
Ontario (Season 1)	0.00018	100.8	10.2	12.92
Ontario (Season 2)	0.0004	102	9.92	12.57
Ontario (Season 3)	0.00022	99.4	11.9	16.32
Victoria	0.00009	44.17	8.38	10.62
NSW	0.00014	55.83	8.77	8.68

Spain	0.00004	53.64	8.63	8.74
California	0.0011	80.6	5.32	11.39

Table 6.5 shows the relevant statistical measures for the analysis of error achieved over the entire time series. Large standard deviation is noteworthy in case of Ontario market which experiences large values of standard deviation of error in all seasons of the year. Moreover the forecasts in Ontario market exhibit highest value of maximum error which testifies the presence of highest degree of volatility in this market.

6.9 Summary

A hybrid intelligent model was proposed for improving the prediction accuracy in the task of electricity price time series forecasting. The proposed model hybridizes Recurrent Neural Network with FHN excitable system of coupled equations for price forecasting. The forecasting performance in stable regions and spiking regions was found to be satisfactory, over many deregulated markets in the world over different seasons, verifying its adaptability.

Success of this synergistic combination of RNN and coupled system of equations presents exciting opportunities for future work in day-ahead prediction in this time series system using multi-scale system. If the exact underlying slow-fast interaction mechanism can be extracted from time series, results can be considerably improved. Exact multiscale modelling will be the future approach in this context.

Chapter 7

Multi-Scale Modelling of Electricity Price Time Series using Multi-Scale Neural Network

In this chapter spikes are seen as critical transitions or tipping points which are sudden changes in the dynamical system. Typical examples of critical transition are the changes occurring in climate, medicine, ecological systems and finance. Although several approaches have been proposed for analysis of specific models, a detailed deterministic model has not been developed for exact modelling of spikes in time series. There exist strand of models where early warning signals for these tipping points have been developed. In this work a mathematical framework is adopted to obtain the qualitative and quantitative understanding of critical transition to check the underlying nonlinear scenarios.

A variety of systems showing the behaviour of nonperiodic spiking can be mathematically described by means of slow and fast systems coupled together [136-137]. In a two dimensional system, nonperiodic spiking appears when the system state is in the vicinity of fixed point or limit cycle bifurcations (subcritical/supercritical, andronov saddle-node collisions). On the other hand higher dimensional systems can exhibit more varied and complex behaviour such as bursting [138], mixed mode oscillations [139] etc. In the case of electricity price time series, which occupies a high dimensional phase space, irregular spikes occur as the result of complex bifurcation sequences which have deterministic origin instead of stochastic, which implies that spiking sequences are chaotic instead of irregular. In most of the cases where model equations are available, the phenomenon has been understood in terms of Shilnikov Homoclinic Chaos (SHC) [140]. SHC occurs in a typical three dimensional phase space when a growing periodic orbit reaches a saddle focus and become biasymptotic to it, i.e they become homoclinic. Typical spiking time series consists of large amplitude

pulses (corresponding to homoclinic orbit in phase-space), while the interspike time intervals consist of small amplitude chaotic oscillations. Although, this mechanism is the one suggested by Shilnikov theorem, which predicts the occurrence of complex dynamics near homoclinicity whenever the saddle-focus, with eigenvalues in the linearized neighbourhood $(\mu, -\rho \pm i\varpi)$ where $(\mu, \rho) > 0$ and they satisfy the condition $|\rho / \mu| < 1$. However, SHC is not a generalized explanation for the presence of chaotic spiking, though each spike is definitely associated with a reinjection mechanism. For example, and in a particular slow-fast system, a Hopf bifurcation is sometimes followed with a period doubling cascade which generates a sequence of small-periodic and chaotic excitable dynamics, that develops before relaxation oscillations take place. As the amplitude of the chaotic attractors grows, some particular fluctuations of the occurring chaotic dynamics spontaneously trigger excitable spikes in an erratic but deterministic manner [141]. Such phenomena are referred to as incomplete homoclinic scenarios [142] since, in the appropriate parameter range, mimic trajectories close to Shilnikov conditions.

7.1 Slow-Fast Systems

A detailed introduction to the role of fast-slow dynamical systems for critical transitions has been developed in [143]; in particular, the link between applications and a rigorous mathematical theory has been established. In this section we recall the major definitions and results from multiple time scale dynamics [144-145] that are required to define critical transitions. A fast-slow system of ordinary differential equations (ODEs) is given by

$$\begin{aligned} \varepsilon \dot{a} &= f(a, b, \varepsilon), \\ \dot{b} &= g(a, b, \varepsilon), \end{aligned} \tag{7.1}$$

where $0 < \varepsilon << 1$, $a \in R^n$ are fast variables and $b \in R^m$ are slow variables.

Changing in (7.1) from the slow time scale s to the fast time scale $\tau = t/\varepsilon$ gives

$$\begin{aligned} a' &= f(a,b,\varepsilon), \\ b' &= \varepsilon g(a,b,\varepsilon), \end{aligned} \tag{7.2}$$

Now if we consider the singular limit $\varepsilon \to 0$, we obtain the slow subsystem

$$\begin{aligned} 0 &= f(a,b,\varepsilon), \\ \dot{b} &= g(a,b,\varepsilon), \end{aligned} \tag{7.3}$$

which is a differential-algebraic equation restricted to the critical manifold

$$C_0 = \{(a,b) \in R^{m+n} : f(a,b,0) = 0\}$$

The slow subsystem is obtained as the singular limit of (7.2)

$$\begin{aligned} a' &= f(a,b,\varepsilon), \\ \dot{b} &= 0, \end{aligned} \tag{7.4}$$

here the slow variables can be viewed as parameters. The flows generated by (7.3) and (7.4) are called the slow flow and the fast flow respectively. Assume a point $p \in C_0$ which is an equilibrium point of the fast subsystem. The critical manifold is normally hyperbolic at $p \in C_0$ is the *n×n* matrix $D_a f(p)$ is a hyperbolic i.e. has no eigenvalues with zero real parts. In this case, the implicit function theorem provides a map $h_0 : R^m \to R^n$ that describes C_0, in the neighbourhood of *p*, as a graph

$$C_0 = \{(a,b) \in R^{m+n} : a = h_0(b)\}. \tag{7.5}$$

Then the slow subsystem (7.4) can be written more concisely as

$$\dot{b} = g(h_0(b),b). \tag{7.6}$$

If all eigenvalues of $D_x f(p)$ are positive(negative) then C_0 is repelling (attracting), otherwise it is of saddle-type. Observe that C_0 is attracting at p if and only if the fast subsystem has a stable hyperbolic equilibrium at p.

Fenichel's Theorem provides a complete description of the dynamics for normally hyperbolic invariant manifolds for sufficiently smooth vector fields (f, g).

Theorem 7.1 (Fenichel's Theorem) [Neil Fenichel, 1979]: Suppose $S = S_0$ is a compact normally hyperbolic submanifold of the critical manifold C_0. Then for $\varepsilon > 0$ sufficiently small the following holds:

- There exists a locally invariant manifold S_ε diffeomorphic to S_0. Local invariance means that S_ε can have boundaries through which trajectories enter or leave.
- The flow on S_ε converges to the slow flow as $\varepsilon \to 0$.
- S_ε is normally hyperbolic and has the same stability properties with respect to the fast variables as S_0 (attracting, repelling or saddle type).
- S_ε is usually not unique. In regions that remain at a fixed distance from the boundary of S_ε.

All manifolds satisfying above four conditions lie at a Hausdorff distance $O(e^{-K/\varepsilon})$ from each other for some $K > 0$ with $K = O(1)$.

It should be noted that in the above theorem is stated for slow-fast systems where the invariant manifold corresponding to unperturbed system dynamics is considered as the slow manifold and the critical transitions are considered as variation in fast manifold. This definition is different from the notion of slow-fast systems used in this thesis where the unperturbed invariant manifold of dynamics is considered to be the fast manifold. Later in

this chapter, this theorem is employed (in theorem 7.2), however, using the notion of slow-fast systems used in this thesis.

A wide class of nonlinear physical systems contains slow and fast dynamic processes that occur at different moments. Recent research results show that neural networks are very effective for modelling the complex nonlinear systems with different time scales when we have incomplete model information, or even when we consider the plant as a black-box [146]. Dynamic neural networks with different time scales can model the dynamics of the short-term memory of neural activity levels and the long-term memory and the dynamics of unsupervised synaptic modifications [147]. Different methods have been applied in this domain. The stability of equilibrium of competitive neural network with short and long-term memory was analyzed in [148] by a quadratic-type Lyapunov function. Refs. [149-151] presented new methods of analyzing the dynamics of a system with different time scales based on the theory off low invariance. The K-monotone system theory was used for analyzing the dynamics of a competitive neural system with different time scales in [152]. The past decade has witnessed great activities in stability analysis, identification and control with continuous time dynamic multi-time scale neural networks [153]. In [154], Sandoval et al. developed new stability conditions using the Lyapunov function and singularly perturbed technique

7.2 Multi-Scale Neural Network (MSNN)

Numerous systems in the industrial fields demonstrate non-linearities and uncertainties, which can be considered as partially or totally black-box. Dynamic neural networks have been applied in system identification and control for those systems for many years and due to the fast adaptation and superb learning capability, they have transcendent advantages

compared to the traditional neural network methods [146, 155]. Dynamic neural networks with different time-scales can model the dynamics of the short-term memory and the long-term memory. The dynamics of dynamic neural networks with different time-scales are extremely complex, exhibiting convergence point attractors and periodic attractors [147]. Networks where both short-term and long-term memory are dynamic variables cannot be placed in the form of the Cohen-Grossberg equations [156-157]. However, a large class of competitive systems have been identified as being ”generally” convergent to point attractors even though no lyapunov functions have been found for their flows.

Until now some artificial neural networks based approaches have been proposed for system identification. However these models have not been applied for function approximation tasks of spiky time series modelling.

7.3 MSNN for Electricity Price Modelling

In this work, we consider the electricity price time series as nonlinear system with multiple time scales and model it via a discrete time dynamic neural network with different time scales including both fast and slow phenomena. Two RNN identifiers at different scale, based model are proposed for nonlinear systems identification. For two RNN identifiers, the steepest descent and singularly perturbed techniques are used to develop the update laws for both dynamic neural network weights.

In this section we consider the problem of identifying the nonlinear system namely electricity price time series which is assumed to belong to class of singular perturbation nonlinear systems with two different time scales described by equation (7.2). In order to identify the nonlinear dynamic system, we employ the MSNN with two time scales:

$$x_{nn}(t+1) = V_1 \cdot f(U_1 \cdot X_{nn} + W_1 \cdot x_{nn}(t) + D_1 y_{nn}(t))$$
$$y_{nn}(t+1) = y_{nn}(t) + \varepsilon\big(V_2 \cdot f(U_2 \cdot Y_{nn} + W_2 \cdot y_{nn}(t) + D_2 x_{nn}(t))\big) \tag{7.7}$$

Here $x_{nn} \in R^{n_1}$ and $y_{nn} \in R^{n_2}$ are fast and slow variables, $U_i \in R^{n_{i1} \times k_{i1}}$ (i=1,2) is the weight matrix connecting input layer to hidden layer for fast/slow subsystem for i=1/2, G_i is the weight matrix connecting slow/fast variables to hidden layer of fast/slow subsystem for i=1/2. V_i is the weight vector connecting hidden layer of slow/fast variables to output layer of fast/slow subsystem for i=1/2. W_i is the feedback weight connecting output of fast/slow variables to hidden layer of fast/slow subsystem for i=1/2. $f(Z)$ ($Z = \begin{bmatrix} z_1 & z_2 & . & . & z_k \end{bmatrix}^T$ is the vector of squashing function (in this work we use sigmoid) = $\begin{bmatrix} \sigma(z_1) & \sigma(z_2) & . & . & \sigma(z_k) \end{bmatrix}^T$

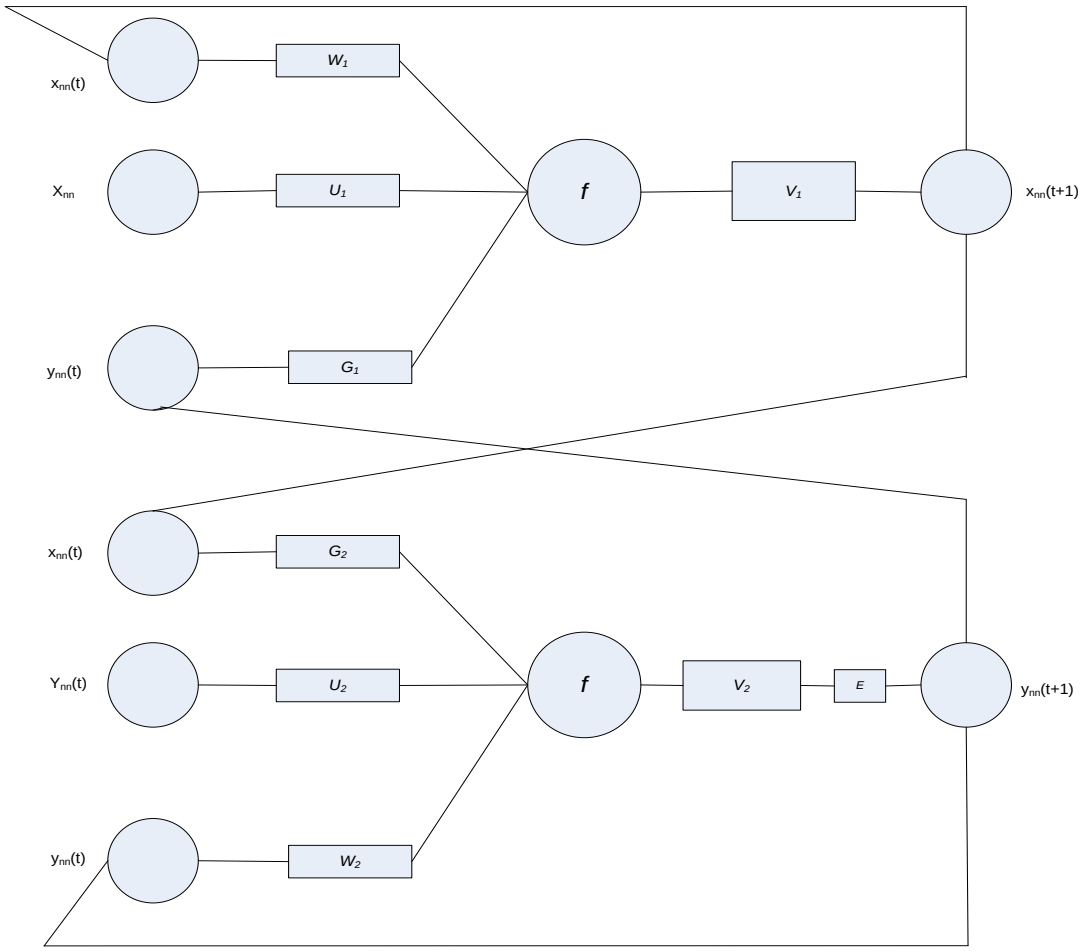


Fig. 7.1.Multiple Scale Neural Network

In general the multiple scale neural network is trained to follow the states of slow and fast subsystems. However in this context only time series values are available. In this work we use delay embedding based vector as the input X_{nn} to the fast subsystem. The slow manifold or the slow subsystem is responsible for critical transitions. Or in other words, perturbations in the slow variable cause the so called tipping points such as spiking. There are several attributes that have been observed in systems before critical transitions such as

- There is a slowing down of system states as system recovers from slowly from perturbation.
- There is increase in system variance as critical transition is approached.
- The autocorrelation of time series increases before transition.

In this work we embed these observations into mathematical modelling of system dynamics. In particular, system specific criteria can be introduced for learning in MSNN, however in this case due to complex nature of time series and difficulty in extraction of system specific dynamics, we have employed generic indicator which are applicable to large class of transitions. The local variance of time series and autocorrelation are used as inputs to the slow subsystem. These two networks are coupled using coupling matrix G_i. The presence of multiple dimensions in fast and slow subsystems allows for possible dynamics in phase space due to complex interaction of slow and fast manifolds and nature of eigenvalues at the equilibrium.

7.4 MSRNN learning

The learning of multiple scale neural network can be an intricate issue. In this work we employ steepest descent and geometrical singular perturbation theory for modelling. Firstly there is no way of learning the scaling parameter ε. Secondly the slow and fast subsystems represent two different manifolds which can be difficult to learn using single time series. The fact that the two submanifolds evolve at different time scales, it is not possible to train them together with same learning algorithms.

In this work we propose a different schema for learning of both fast and slow neural networks. The parameter ε describes the separation of time scales. These different scales

are characteristic of slow fast systems which are an important class of singular perturbation system. One successful strategy to understand fast-slow systems from singular perturbation theory point of view is to understand the case ε = 0 and using this knowledge to analyze perturbation results for ε > 0 sufficiently small; many geometric and asymptotic methods follow this pattern. The full system is trained following the procedure mentioned below

1. *Approximate Learning of fast subsystem*

In the nonspiking regime, given very small value of 0< ε <<1, we assume that ε =0. For ε =0, equation (7.2) results in

$$\begin{aligned} x_{nn}(t+1) &= V_1 \cdot f(U_1 \cdot X_{nn} + W_1 \cdot x_{nn}(t) + D_1 y_{nn}(t)) \\ y_{nn}(t+1) &= y_{nn}(t) + 0 \end{aligned} \tag{7.8}$$

which implies that the slow subsystem is virtually stagnant and there is no variation in slow parameter value which isolates the fast component of MSNN. This observation follows that in the nonspiking regime, where the slow component hardly varies, fast component dynamics are the only functioning mechanism. Using this information, we train the fast subsystem of MSNN in the nonspiking regime. This constitutes the first step of training, where the trained network learns the invariant manifold. In the next step we suggest a theorem which ascertains that this trained network is the near accurate approximation of the fast subsystem.

Theorem 7.2: A set M , which belongs to invariant fold of a flow corresponding to ε =0, then there exists a manifold M_ε that lies within $O(\varepsilon)$ of M and is diffeomorphic to M . Moreover it is locally invariant under the flow including in ε .

The proof of this theorem follows from the first theorem of Fenichel. Using this theorem it can be stated that the neural network trained for ε =0 is a close approximation for the fast subsystem of MSNN for conditions where perturbation is present, or 0< ε <<1.

We exploit this condition for training fast component of the MSNN. Firstly the spiking and nonspiking regimes of time series are identified. The spiking region can be approximately considered as 10 hours before spike, the spiking region and 10 hours after spike. **The network is trained using steepest descent using the same algorithm suggested in chapter 5 using the fixed point dynamics based initialization and learning of network weights.**

2. *Approximate Learning of slow subsystem*

The slow subsystem, which is chosen to be relatively smaller in size is trained on the spiking regime of time series. During this training procedure, the structure of fast subsystem is kept fixed.

3. *Approximate Learning of scaling parameter*

The scaling parameter which reflects the time scaling of slow and fast variable is denoted by ε in the MSNN. A very close approximate of ε can be obtained using spectrum analysis of the time series. The eigenspectrum of Jacobian matrix in table 4.2 of time series approximately denote the different temporal scales at which system dynamics take place [158]. Presence of two time scales is characterized by two main clusters of eigenvalues. If the absolute value of largest eigenvalue of the small eigenvalue cluster is denoted as λ_s and the smallest eigenavlue of the large eigenvalue cluster is denoted by λ_l, then the scaling parameter ε can be obtained as

$$\varepsilon = \frac{\lambda_s}{\lambda_l} \tag{7.9}$$

4. *Learning of full MSRNN*

The above 3 steps provide us the network trained with slow and fast component trained individually. In the next step the fully connected MSRNN is trained over the complete time series using steepest descent.

7.5 Summary

In this work, a Multilayer Recurrent Neural Network (MSRNN) has been proposed for improving the prediction accuracy in the task of electricity price time series forecasting. The proposed approach exploits the multi-scale nature of time series and a slow-fast recurrent neural network is proposed. For more accurate modeling of time series and the associated homoclinic scenarios, a multiple scale neural network (MSRNN) is developed. Slow fast systems deal with slow manifold and fast manifold where the key dynamics of time series occur on fast invariant manifold while the dynamics occurring on slow manifold is responsible for intermittent critical transitions. The developed model is trained using singular perturbation theory for slow-fast systems combined with gradient descent algorithm.

The forecasting performance of this model in the spiking and nonspiking region was found to be excellent (implementation results given in chapter 8).

Success of this MSRNN model presents exciting opportunities for future work in this time series system. Further work will include extensive investigation of slow-fast scale analysis of electricity price time series and include the results as heuristics in modeling of MSRNN.

Chapter 8

Results and Discussion

In chapter 5,6 and 7, three RNN based approaches have been proposed with the objective of achieving better prediction accuracy in day-ahead electricity price time series system, with particular focus on spiking regions of time series. The motivation behind the first model, PGRNN, is to learn the inherent invariant dynamics of time series system in order to obtain a better nonlinear modelling. The second model, RNNFHN incorporates multiscale behaviour via employing FHN self-coupled equation system. However due to hybridization, stochastic learning (ES) had to be employed in this model which results in low prediction accuracy in some cases. The motivation behind development of third model was to overcome this limitation and a multi-scale neural network was developed using singular perturbation+gradient descent based deterministic learning approach.

In this thesis these models have been tested on worldwide deregulated markets for various seasons. The details of dataset used are mentioned in next section.

8.1 Data and Preliminary Statistical Analysis

8.1.1 Data

The electricity markets most affected by restructuring and which are studied in used for forecasting include California and PJM markets in the US, Spain, Ontario market in Canada, and NEMMCO market in Australia. In some of these markets, higher energy prices have enticed new generating companies to enter, resulting in a greater competition and higher volatility [159], while in markets like Ontario, huge networks with neighbour markets

and regulated utilities significantly complicates the competition scenario. The implementation results in these five markets have been presented. The trained model was implemented on unseen testing datasets given in Table 8.1. The **training dataset** includes one month data prior to testing period. The datasets used incorporate the defining characteristics of electricity price time series, like high volatility, occasional extreme jumps, and volatility clustering.

Table 8.1 presents a data description of the electricity price series in our study. For the California and PJM, price data are based on Dow Jones Electricity Price Indexes and are drawn from Datastream International. These are volume-weighted price indexes based on Dow Jones Firm On-Peak daily indexes and are denominated in U.S. dollars per MW. For PJM, the data from PECO zone is considered for analysis.

In this proposed work, six deregulated markets worldwide were investigated and experimented upon. The idea of deregulation has disseminated among many countries with large power markets referable to its lowering in price effect. The markets most affected by restructuring include California, PJM, UK electricity market, Ontario and NEMMCO Australia. The market data was obtained from www.elexon.co.uk for UK market, from www.ieso.ca for Canadian Ontario market, from www.nemmco.com.au for NEMMCO Australian data etc. Some of the literatures have tried to analyze these markets and the effects of deregulation upon them using various volatility indices. The Australian data is obtained from NEMMCO (www.nemmco.com.au) and represents hourly prices. The prices are denominated in Australian dollars per MW. The two trading hubs in the transmission network for which we obtain data are New South Wales (NSW) and Victoria (VIC).

Table 8.1
Testing Dataset Used. Training Set is One Month Data Prior to Testing Dataset.

Market	Season	Testing Data
Ontario(week 1-2)	Spring	Apr 26 – May 9, 2004
Ontario(week 3-4)	Summer	Jul 26 – Aug 8, 2004
Ontario(week 5-6)	Winter	Dec 13 – Dec 26, 2004
NEMMCO Victoria (Week1)	Summer	Jul 26 – Aug 1, 2003
NEMMCO Victoria (Week2)	Summer	Aug 19-Aug25, 2003
NEMMCO Victoria (Week3)	Winter	Dec 1-7 , 2003
NEMMCO NSW(Week1)	winter	Jan 22 – Jan 28,2006
NEMMCO NSW(Week2)	Fall	May 21 – May 27,2006
NEMMCO NSW(Week3)	Summer	Aug 20 – Aug 26, 2006
NEMMCO NSW(Week4)	Spring	Oct 22 – Oct 28, 2006
Spain	Summer	Aug 19 – Aug 25, 2002
California	Spring	Feb 21 – Feb 28, 2000
PJM (week 1)	Spring	Feb 22- Feb 28, 2006
PJM (week 2)	Fall	May 17-May 23, 2006
PJM (week 3)	Summer	Aug 23-Aug28, 2006
PJM (week 4)	Winter	Nov 22-Nov 28, 2006

8.1.2 Summary Statistics

Comparisons across the price levels show notable differences in volatility among the markets. For example, Ontario prices exhibit more variability than either the U.S. or

Australian prices. In all three sets of price figures, the volatility is dominated by episodes of extremely high prices. During the California power crisis of summer 1999 prices spiked to over $1,500 per MW. The Australian prices for NSW and Vic show a greater frequency of extreme price spikes than any of the other series. The time series are more clearly mean-reverting. There are episodes of "clustering" or volatility that are visibly more frequent than just the price spikes that dominate volatility in the price levels series.

Table 8.2
Statistics of Training Data (The units of Mean, Std. Dev, Min and Max are Euro/MW for Spain and $/MW for rest of the markets)

	PJM	Ontario	Spain	Victoria	NSW	California
Mean	38.65	46.9	3.81	42.32	38.63	26.34
Std. Dev	13.3	18.8	1.469	170	60.97	17.1
Minimum	21.18	10	1.37	11	14.35	9.97
Maximum	70.16	94.8	7.4	2542	529.8	55.32
Skewness	57.2	51.47	37.93	12.5	74.75	35.2
Kurtosis	2.11	3.22	2.0	179.24	32.4	1.64

Table 8.2 presents summary statistics for the time series for each of the six markets. We report the mean, standard deviation, minimum, maximum, skewness, kurtosis for time series. The hourly standard deviations of the time series are, on average, very high and widely dispersed across trading hub series even within countries. The magnitude of and cross-sectional dispersion in standard deviations is much lower in spain than the U.S. and much higher in Australia and Ontario than in the U.S. Manifestly California market is the most studied one ascribable to the instability occurred during end of 2000. The price trends show volatile behavior around this period as the electricity market began to reach its limits. Large generation capacity leads to higher prices alluring more market participants and thus

altering the competition scenario. On the contrary, the reason behind high volatility in PJM market is the lack of sufficient generation imparting emergency conditions. The Ontario power network is connected to the New York and Midwest electricity markets directly, and to the New England and PJM markets indirectly. It is also connected to the regulated utilities in Quebec and Manitoba, both having significant energy transactions with other utilities in the United States. Volatility of the Ontario electricity market prices is shown to be significantly higher than price volatility in the New England, New York and PJM electricity markets, as well as in other markets around the world, revealing that Ontario's electricity market prices are among the most volatiles world-wide. Table I explains the characteristic nature of Electricity price time series system. High volatility, intermittent spikes etc are clearly evident in this time series system.

None of the series exhibit much skewness, but all have high positive excess kurtosis, which is again likely to signify fat-tailedness due to very large outliers. The magnitude of the kurtosis statistic, which mostly hover around 10 for the various series, is aligned with the magnitude of the range defined by the maximum and minimum in each series.

Table 4.2 demonstrates the defining characteristics of the time series data used from the electricity markets, like high volatility, occasional extreme jumps, volatility clustering etc, Standard deviation is the measure of high volatility index and the high difference between maximum and mean value tells us the time series observes intermittent extreme jumps. The application of the proposed model on these various training and testing data are presented in next section and the relevant analysis is given.

The volatility and non-stationarity in these markets correspond to prominent and sporadic movements of price from the mean value. In the past literature, this behavior was explained based on existence of different regimes of dynamics in phase space at different time which

corresponds to different behaviors in electricity price time series system. Mainly two regimes can be identified in this nonlinear system and the prices transit from one regime to other based on dynamics of the market during that period. During stable periods, prices exhibit mean-reverting behavior with random fluctuations around some long-term mean while in unstable regions fluctuations are more pronounced, and prices experience jumps and short-lived spikes, that is, unpredictable and very pronounced upward jumps shortly followed by steep downward moves. Mean reversion is a very important property of power prices also in turbulent periods because it is responsible for strongly reducing prices after a spike has occurred. A similar behavior is rarely found in financial markets or in commodities markets, [160].

As discussed here, the different deregulated markets exhibit varying dynamic behavior. Although they share one attribute of spiking at irregular intervals with chaotic behaviour in interspiking region, the volatility characteristics in spiking region and chaotic region are different for different markets. As a result, the three models proposed in previous three chapters are expected to perform differently for these markets, which was the observation upon implementation. The details are given in section 8.3 after brief discussion of used performance indices in next section.

8.2 Forecasting Indices Used

Several measurements are used to examine the accuracy of forecast results. The mean absolute Percentage error (MAPE) index is considered here to evaluate the performance of forecast results. MAPE represents the absolute average prediction error between predictions and actual targets, has been chosen as the error metric due to its wide application in evaluating forecasting models. MAPE is defined as:

:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|P_i - \overline{P}_i\right|}{\overline{P}_i} \tag{8.1}$$

Here N is the total number of data, $\overline{P}_i$ is the actual price and P_i is the predicted price. Many texts use variants of MAPE, namely mean absolute daily error (MADE), weekly mean absolute percentage error (MAWE) etc. as forecasting error indices as low prices can have adverse shooting effects on MAPE. In order to avoid the adverse effect of very small prices, sometimes WMAPE is adopted and compared with those in the literature.

$$WMAPE = \frac{1}{168}\sum_{i=1}^{168}\frac{\left|P_i - \overline{P}_i\right|}{\overline{P}^{168}}, \text{ where} \tag{8.2}$$

$$\overline{P}^{168} = \frac{1}{168}\sum_{i=1}^{168}\overline{P}_i$$

However for fair comparison with other benchmark papers mentioned in later chapters, we use both MAPE and WMAPE in this work as a performance index for the corresponding markets. For California, Ontario, Victoria and NSW day-ahead markets, MAPE has been used, while for PJM market WMAPE has been used for performance evaluation and comparison. For hourly forecasts, MAPE has been used.

8.3 PGRNN Implementation Results

As mentioned in chapter 5, PGRNN model was developed gradually with 1) fixed point based weight initialization (RNN1), 2) fixed point based weight initialization + enforcing fixed point dynamics during learning (RNN2) and 3) fixed point based weight initialization + enforcing fixed point dynamics during learning + local jacobian learning along time series (RNN3). The

individual effects of all these approaches were observed on all markets. Time series data from each market was divided into training and testing sets mentioned in section 8.1. The results are given in table 8.3.

Table 8.3
Prediction Error for Worldwide Deregulated Markets Using Three Proposed Approaches of RNN Learning (Error is in **MAPE** for Ontario and NEMMCO and **WMAPE** for PJM)

Market	**Season**	**RNN 1 (%)**	**RNN 2 (%)**	**RNN 3 (%)**
Ontario(week 1-2)	Spring	13.01	13.12	**12.2**
Ontario(week 3-4)	Summer	13.92	13.82	**12.96**
Ontario(week 5-6)	Winter	12.87	12.99	**12.75**
NEMMCO Victoria (Week1)	Summer	**14.41**	16.21	14.61
NEMMCO Victoria (Week2)	Summer	**14.04**	15.52	14.99
NEMMCO Victoria (Week3)	Winter	**15.38**	16.0	15.7
PJM (week 1)	Spring	6.92	6.9	**6.29**
PJM (week 2)	Fall	5.90	5.36	**5.34**
PJM (week 3)	Summer	5.97	**5.88**	5.9
PJM (week 4)	Winter	7.01	6.91	**6.89**

As observed in table 8.3, the proposed models exhibit interesting results on tested markets. The RNN model trained using learning algorithm employing weight initialization with projected descent and local jacobian gives best results for Ontario markets and PJM markets, while for Victoria market, RNN with simple weight initialization gives best results among the three. This observation can be associated with high spiking nature of time series in Victoria market. Moreover RNN with weight initialization performs better than RNN with Weight Initialization + Projected Descent for most markets except PJM. A possible explanation behind these two observations can be found in the nonlinear theory behind spiking

phenomenon. During learning of RNN models using projected gradient descent based algorithm, the basic assumption is that the system fixed point location and dynamics of system are constant during the course of time series. However it has been suggested in theories and also observed in chapter 6 that spiking is caused due to variation in fixed point behaviour during critical transition associated with spiking. However in case of PJM market, the spiking behaviour is not very prominent and chaotic behaviour dominates the whole time series. Which is why in this scenario RNN trained using all three approaches dominates in PJM market.

In the next table, we present the results corresponding to the RNN trained using all three approaches and compare them with recently published benchmark papers.

Table 8.4
Comparison of Prediction Error for proposed PGRNN trained using combination of all three approaches with Benchmarks (Error is in **MAPE** for Ontario, NSW and NEMMCO and **WMAPE** for PJM)

Market	**Season**	**Benchmark (%)**	**Proposed RNN (%)**
Ontario(week 1-2)	Spring	17.35 [161]	**13.43**
Ontario(week 3-4)	Summer	18.266 [161]	**12.96**
Ontario(week 5-6)	Winter	17.626 [161]	**12.75**
NEMMCO Victoria (Week1)	Summer	22.1 [162]	**12.61**
NEMMCO Victoria (Week2)	Summer	23.1 [162]	**12.99**
NEMMCO Victoria (Week3)	Winter	32.5 [162]	**14.17**
NEMMCO NSW(Week1)	Winter	15.57 [128]	**10.19**
NEMMCO NSW(Week2)	Fall	13.03 [128]	**9.03**
NEMMCO NSW(Week3)	Summer	13.84 [128]	**9.54**
NEMMCO NSW(Week4)	Spring	9.89 [128]	**8.34**

PJM (week 1)	Spring	6.392 [127]	**5.29**
PJM (week 2)	Fall	5.976 [127]	**5.34**
PJM (week 3)	Summer	5.974 [127]	**5.9**
PJM (week 4)	Winter	**6.648 [127]**	**6.49**

It can be observed that the accuracy obtained using the proposed model is better than benchmark models in most of the markets except PJM in season of winter. In case of Ontario and Victoria markets, the improvement in accuracy achieved is considerable. It can be deduced that the proposed model accurately models the time series and the 24 hours ahead market can be forecasted using this approach. The corresponding predicted curves for all markets are shown in figures 8.2 – 8.15 below.

8.3.1 Results for PJM market

The implementation results for PGRNN for PJM market are shown from Fig. 8.1-8.4 where actual and predicted curves are presented.

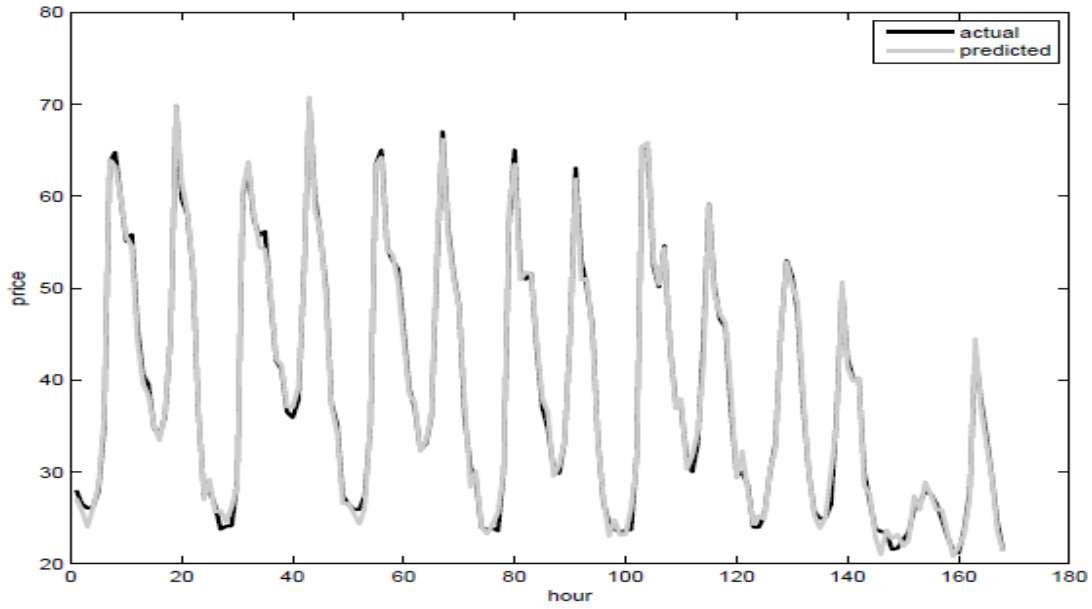


Fig. 8.1. Actual and Predicted curve in PJM market during Week1 for PGRNN

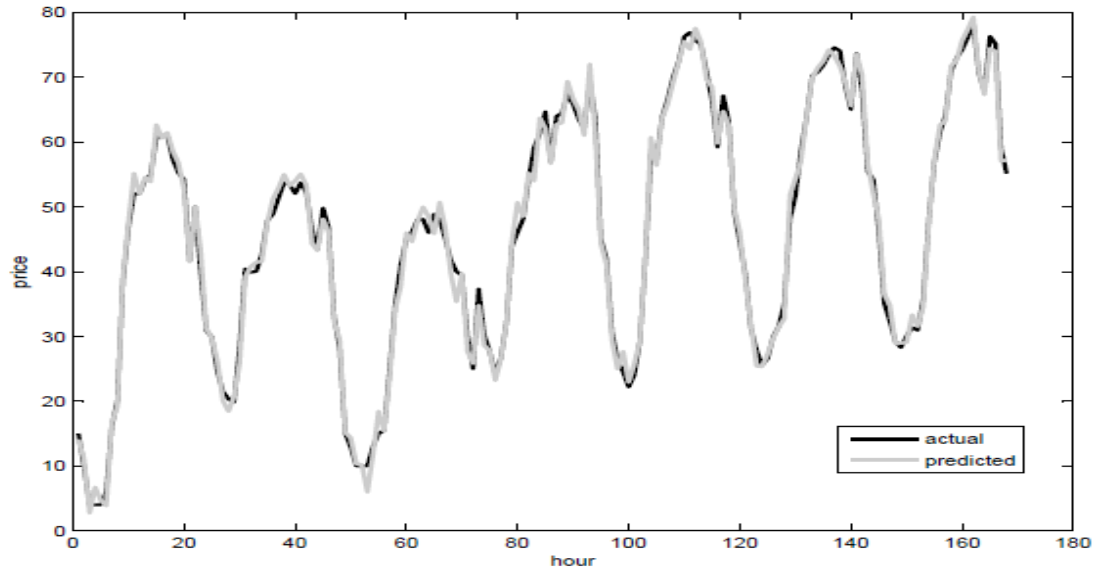


Fig. 8.2. Actual and Predicted curve in PJM market during Week2 for PGRNN

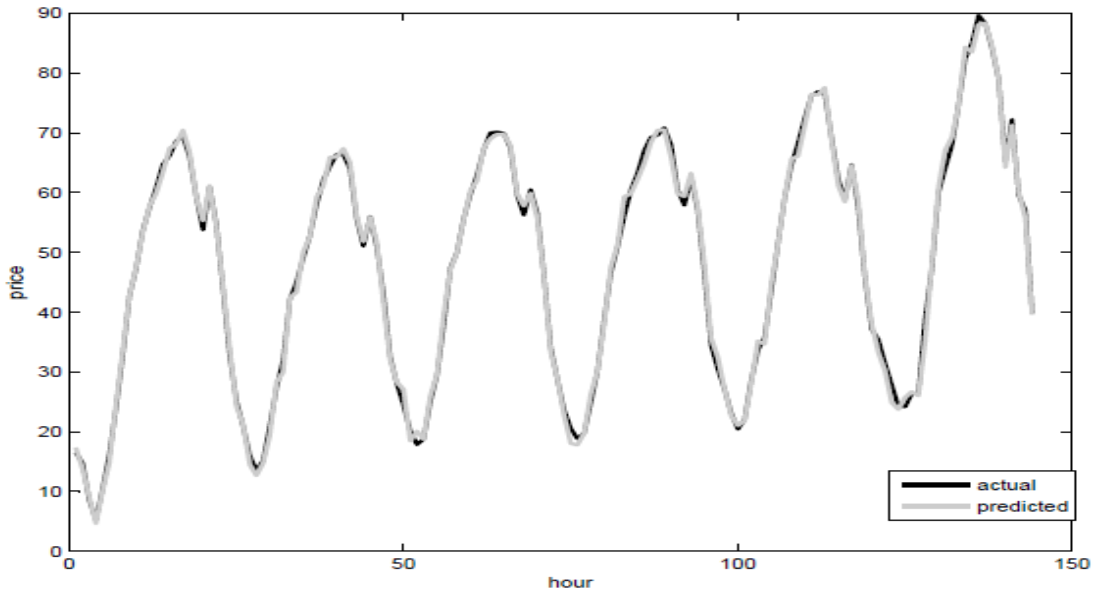


Fig. 8.3. Actual and Predicted curve in PJM market during Week3 for PGRNN

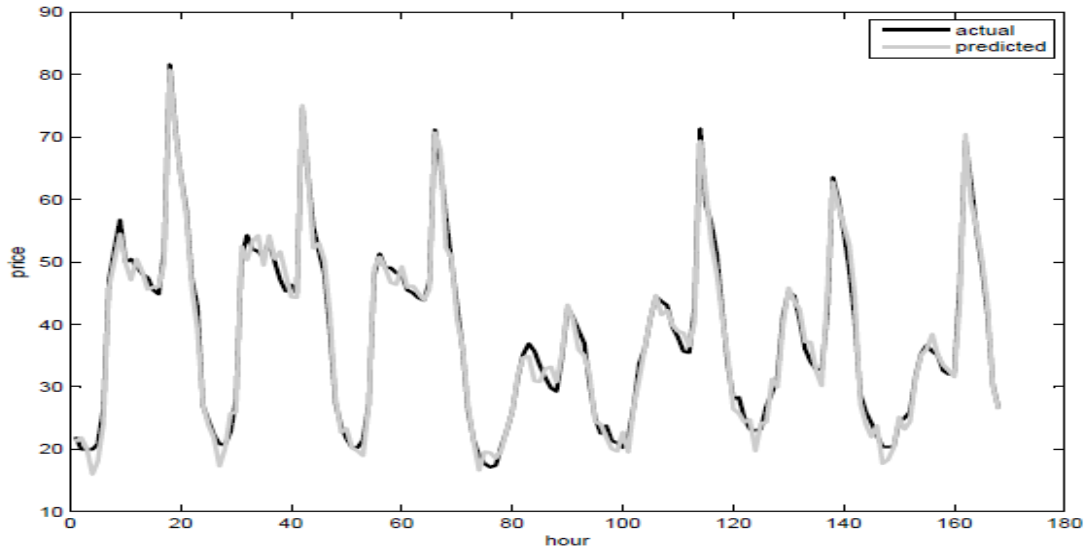


Fig. 8.4. Actual and Predicted curve in PJM market during Week4 for PGRNN

In the above four figures, it can be observed that the testing dataset for PJM market

consists mainly of chaotic oscillations with almost no presence of spikes. The predicted and actual curve comparison confirms that PGRNN works very well in chaotic regions of time series.

8.3.2 Results for Ontario Market

The implementation results for PGRNN for Ontario market are shown from Fig. 8.5-8.7 where actual and predicted curves are presented.

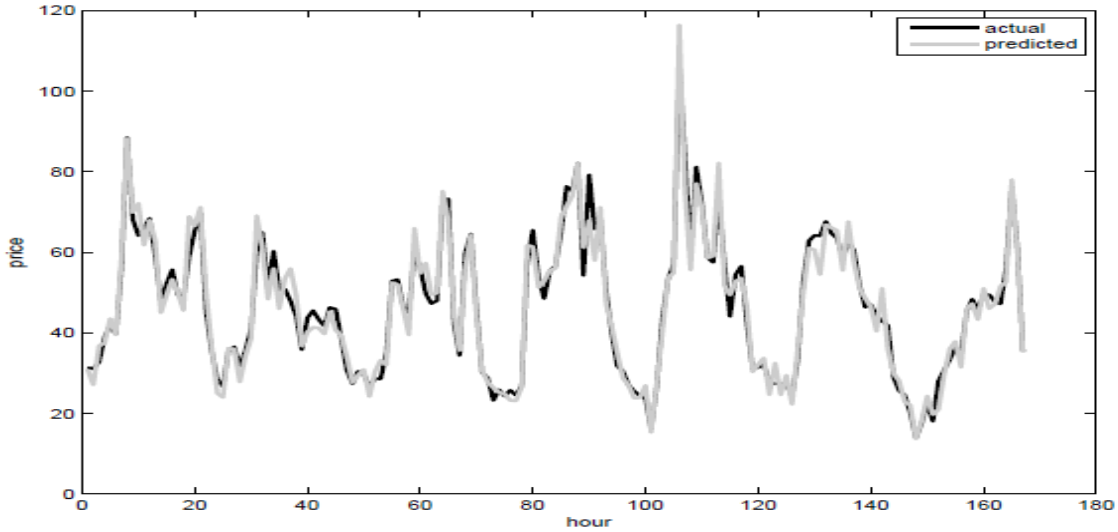


Fig. 8.5 Actual and Predicted curve in Ontario market during Week1 for PGRNN

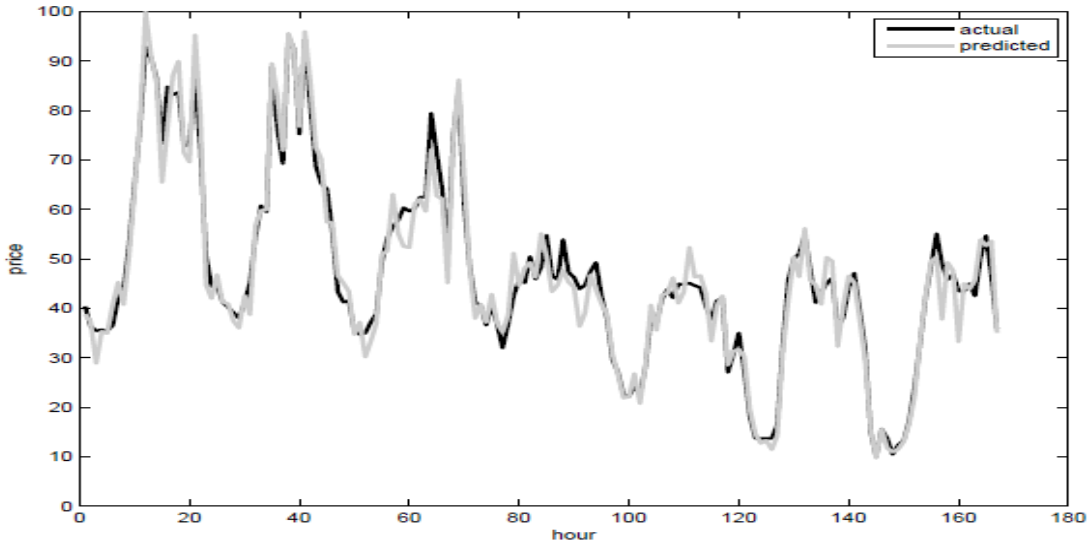


Fig. 8.6. Actual and Predicted curve in Ontario market during Week2 for PGRNN

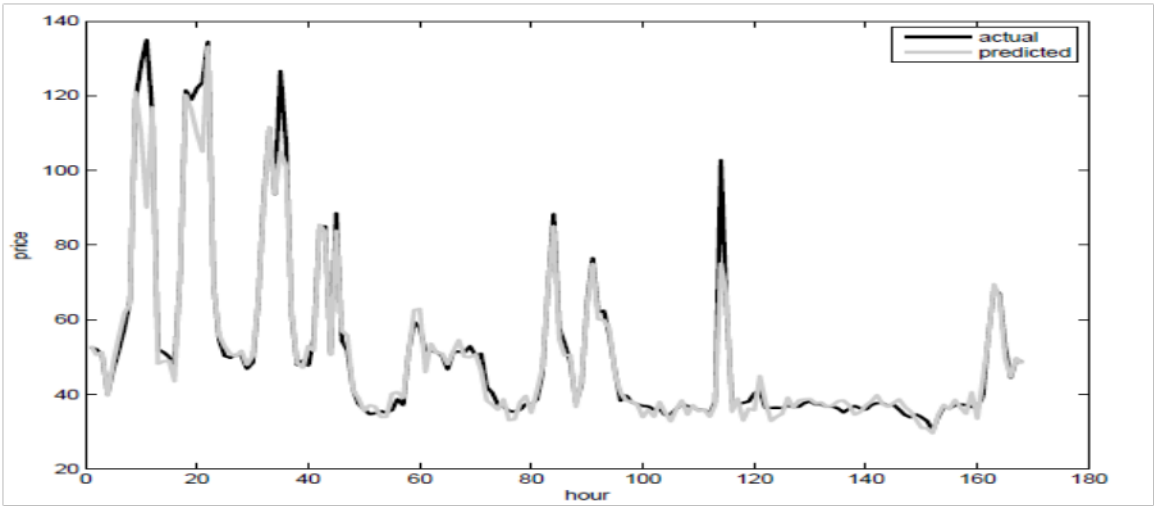


Fig. 8.7 Actual and Predicted curve in Ontario market during Week3 for PGRNN

In the above three figures, it can be observed that the testing dataset for Ontario market consists of highly volatile chaotic oscillations with intermittent presence of single spikes. The developed model performs with considerable accuracy in chaotic regions while

exhibiting few misses during spiking.

8.3.3 Results for Victoria Market

The implementation results for PGRNN for Victoria market are shown from Fig. 8.8-8.10 where actual and predicted curves are presented.

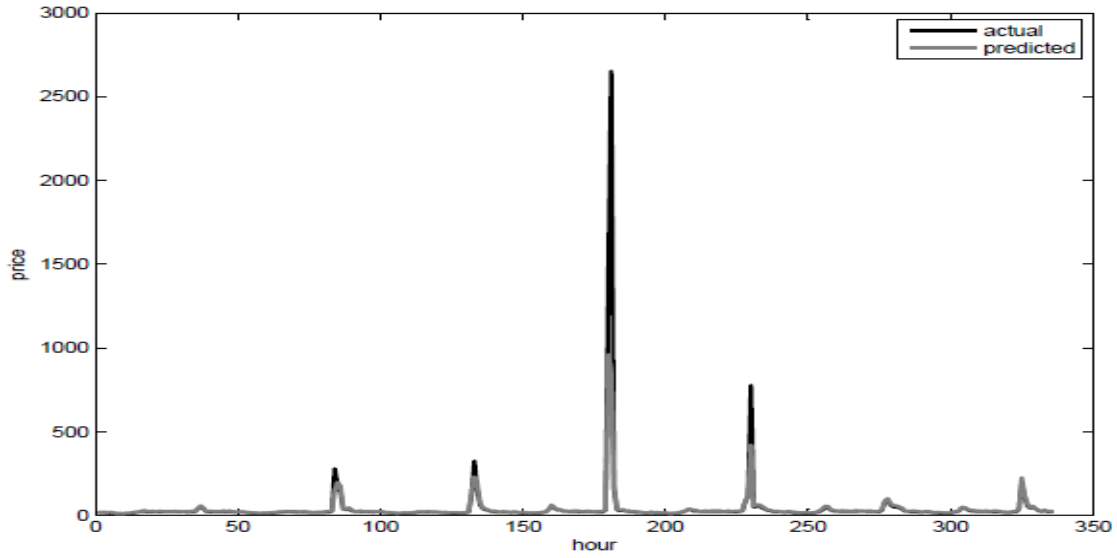


Fig. 8.8 Actual and Predicted curve in Victoria market during Week1 for PGRNN

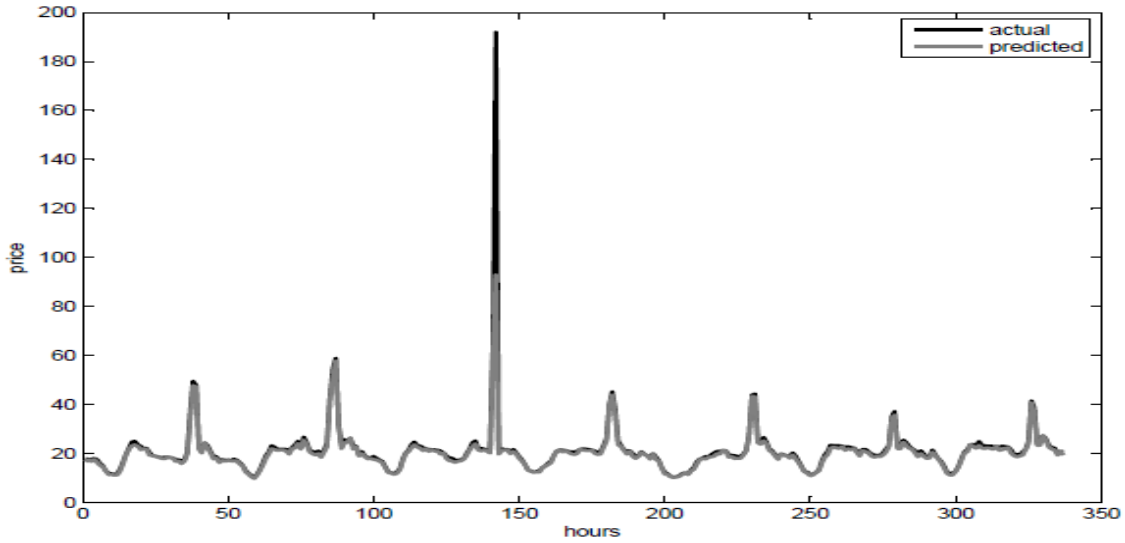


Fig. 8.9 Actual and Predicted curve in Victoria market during Week2 for PGRNN

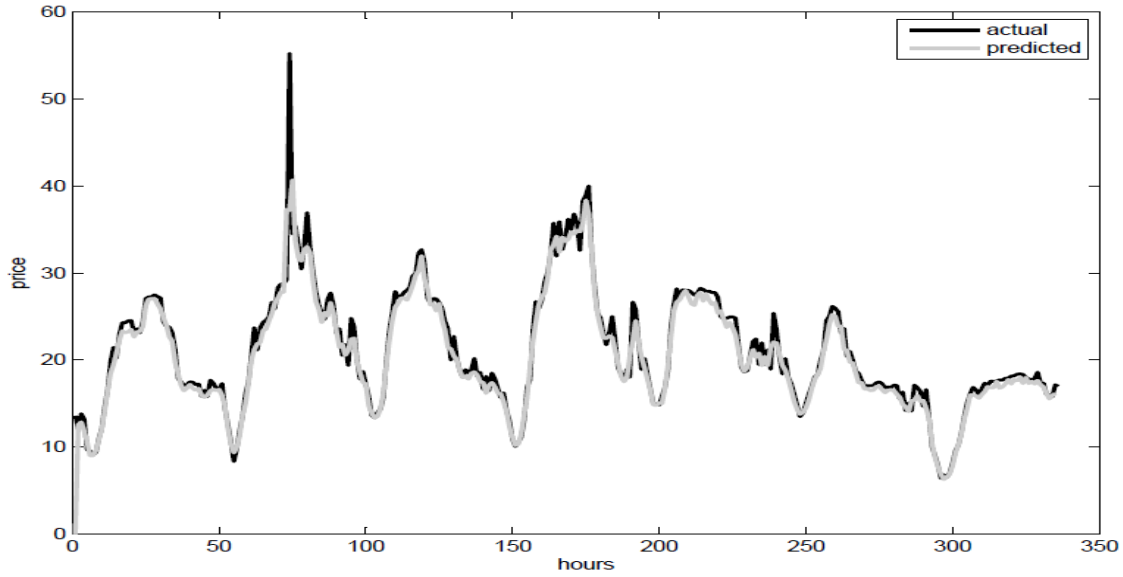


Fig. 8.10 Actual and Predicted curve in Victoria market during Week3 for PGRNN

In the above three figures, it can be observed that the testing dataset for Victoria market consists of low volatile chaos during week 1 and 2 while highly volatile chaotic oscillations in week 3 with intermittent presence of single spikes. The developed model

performs with considerable accuracy in chaotic regions while the accuracy in spiking region is not very satisfactory.

8.3.4 Results for NSW market

The implementation results for PGRNN for NSW market are shown from Fig. 8.11-8.14 where actual and predicted curves are presented.

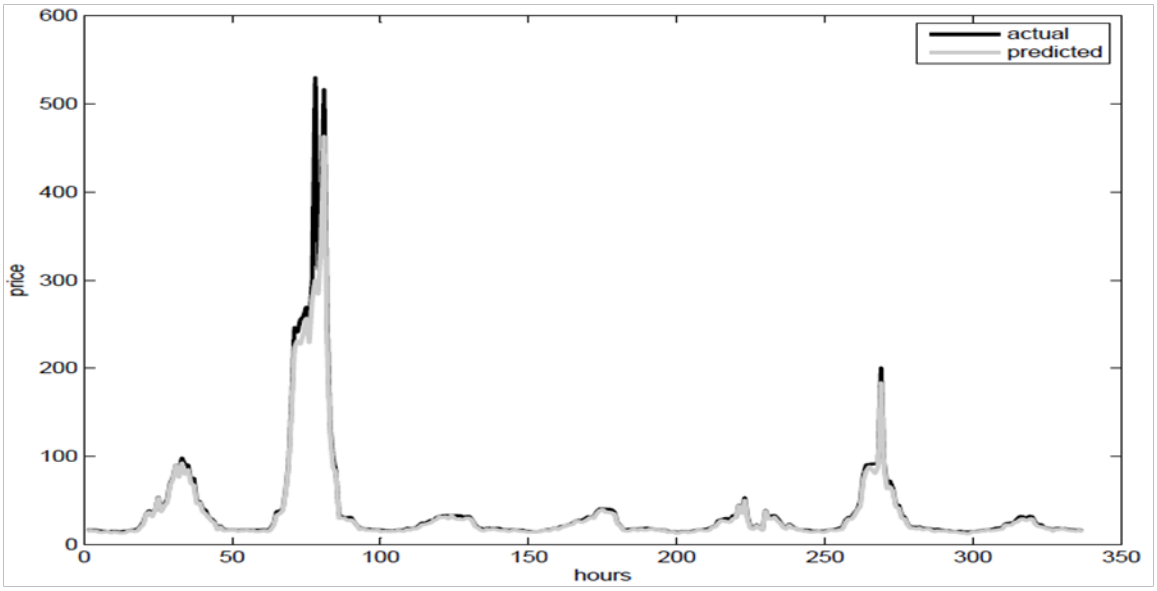


Fig. 8.11 Actual and Predicted curve in NSW market during Week1 for PGRNN

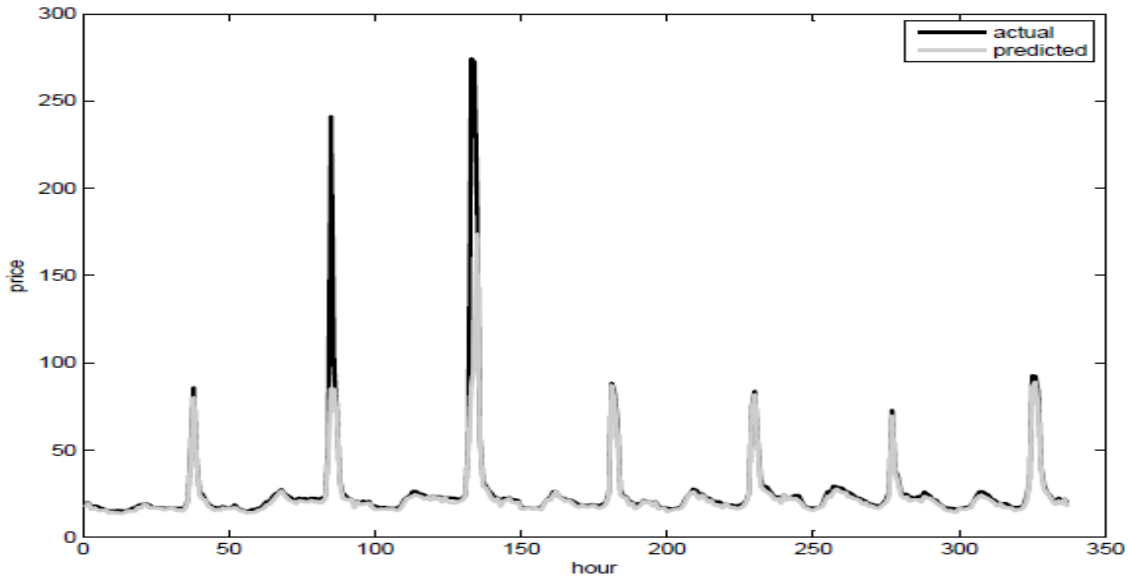


Fig. 8.12. Actual and Predicted curve in NSW market during Week2 for PGRNN

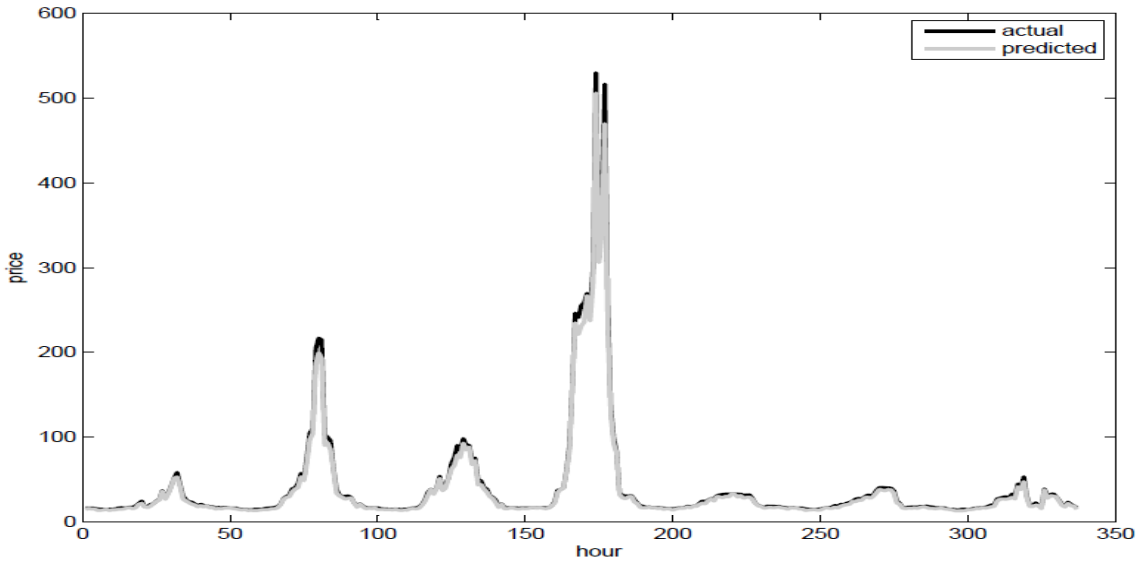


Fig. 8.13. Actual and Predicted curve in NSW market during Week3 for PGRNN

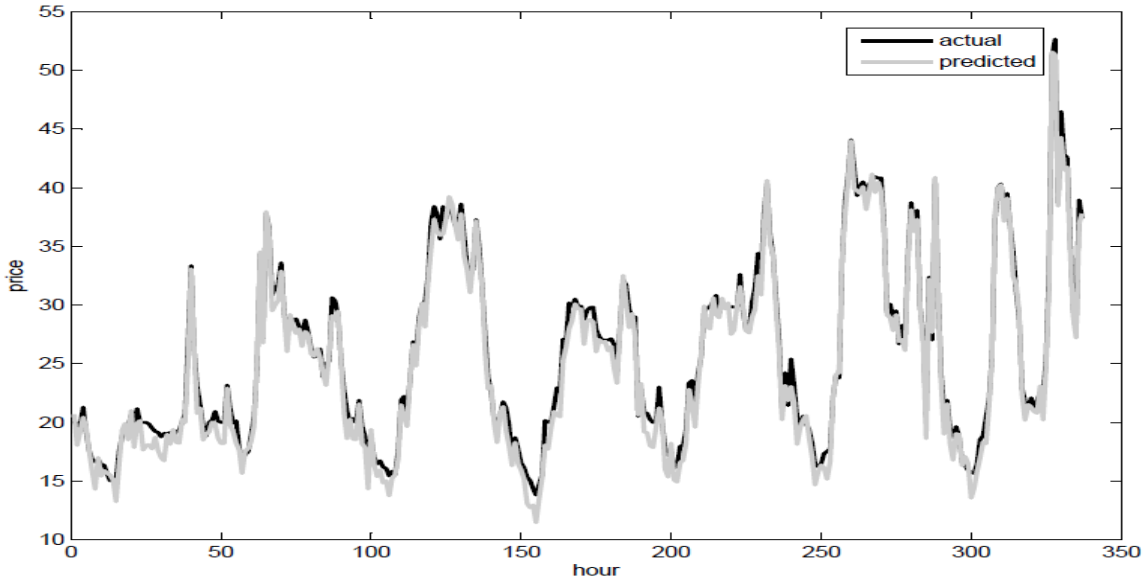


Fig. 8.14. Actual and Predicted curve in NSW market during Week4 for PGRNN

In the above four figures, it can be observed that the testing dataset for NSW market consists of low volatile chaos during week 1, 2 and 3 with intermittent spikes while highly volatile chaotic oscillations in week 4. However the unique feature of week 1 and week 4 is the nature of spikes which is in the form of low frequency bursting. The developed model performs with considerable accuracy in chaotic regions while the accuracy in spiking region is not very satisfactory.

8.4 RNNFHN Implementation Results

The proposed hybrid model was implemented for forecasting of day ahead prices in testing data. Time series data from each market was divided into training and testing sets. Training of hybrid model was carried out in supervised manner and the performance of trained model was examined on unseen test data, without any adaptation. The corresponding performance curves, which show the comparison of predicted and actual results, are shown in Fig 8.15-8.28. The comparison to benchmark results, previously published in literature, is given in table 8.5.

Table 8.5

Comparison of Proposed RNNFHN and Benchmark Approaches for Worldwide Markets (Error is in **MAPE** for Ontario, NSW and NEMMCO and **WMAPE** for PJM)

Market	Season	Benchmark (%)	Proposed RNN (%)
Ontario(week 1-2)	Spring	17.35 [161]	**12.65**
Ontario(week 3-4)	Summer	18.266 [161]	**12.71**
Ontario(week 5-6)	Winter	17.626 [161]	**12.04**
NEMMCO Victoria (Week1)	Summer	22.1 [162]	**13.5**
NEMMCO Victoria (Week2)	Summer	23.1 [162]	**14.02**
NEMMCO Victoria (Week3)	Winter	32.5 [162]	**15.66**
NEMMCO NSW(Week1)	Winter	15.57 [128]	**10.39**
NEMMCO NSW(Week2)	Fall	13.03 [128]	**11.28**
NEMMCO NSW(Week3)	Summer	13.84 [128]	**8.93**
NEMMCO NSW(Week4)	Spring	9.89 [128]	**9.53**
PJM (week 1)	Spring	6.392 [127]	**5.95**
PJM (week 2)	Fall	**5.976 [127]**	5.99
PJM (week 3)	Summer	**5.974 [127]**	6.71
PJM (week 4)	Winter	6.648 [127]	**6.65**

8.4.1 Results for Ontario market

The implementation results for RNNFHN for Ontario market are shown from Fig. 8.15-8.17 where actual and predicted curves are presented.

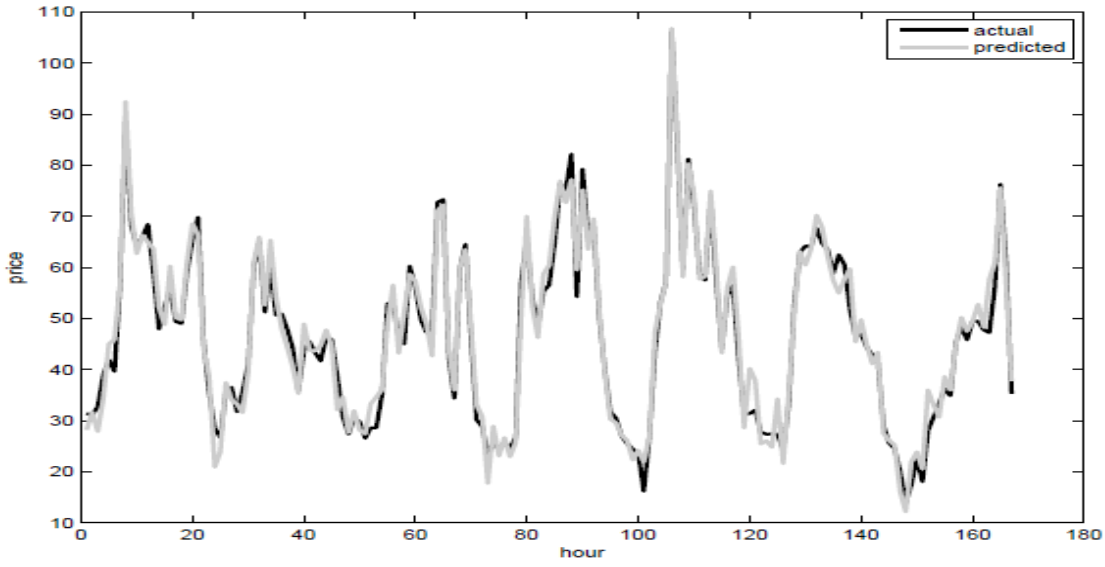


Fig. 8.15. Predicted and Actual Prices for Ontario Market for Week 1 for RNNFHN

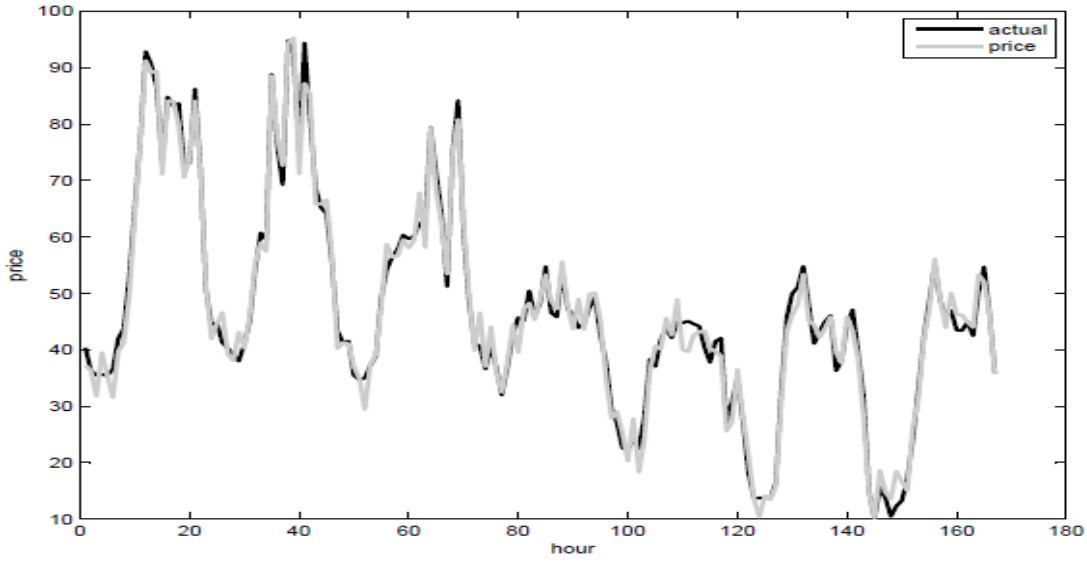


Fig. 8.16.Predicted and Actual Prices for Ontario Market for Week 2 for RNNFHN

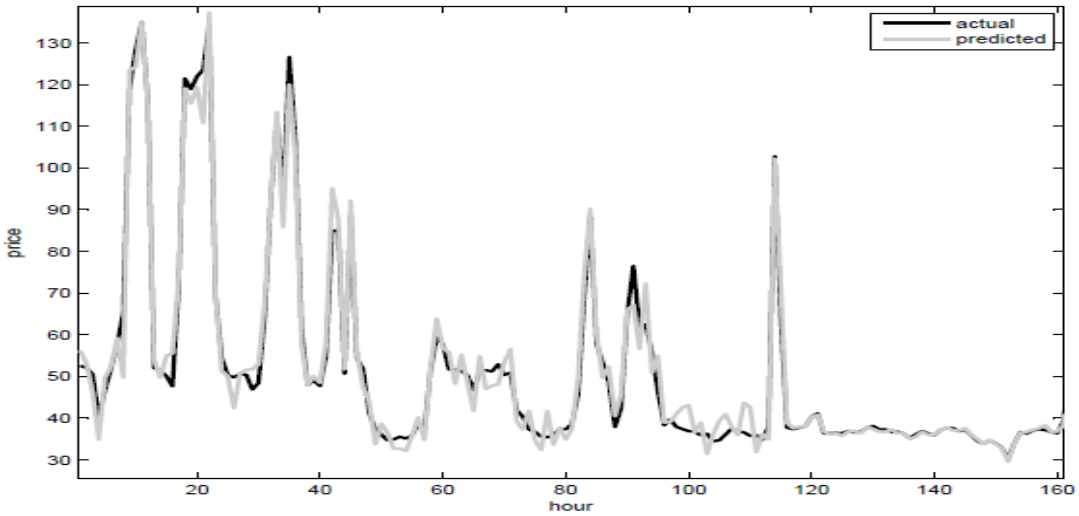


Fig. 8.17 Predicted and Actual Prices for Ontario Market for Week 3 for RNNFHN

In the above three figures, it can be observed that the testing dataset for Ontario market consists of highly volatile chaotic oscillations with intermittent presence of single spikes. The developed model performs well in spiking region with considerable accuracy. However the performance in chaotic region was observed to be rather unsatisfactory, especially when compared with PGRNN model. In chaotic region the behaviour is oscillatory ripples, especially during week 3.

8.4.2 Results for PJM market

The implementation results for PGRNN for PJM market are shown from Fig. 8.18-8.21 where actual and predicted curves are presented.

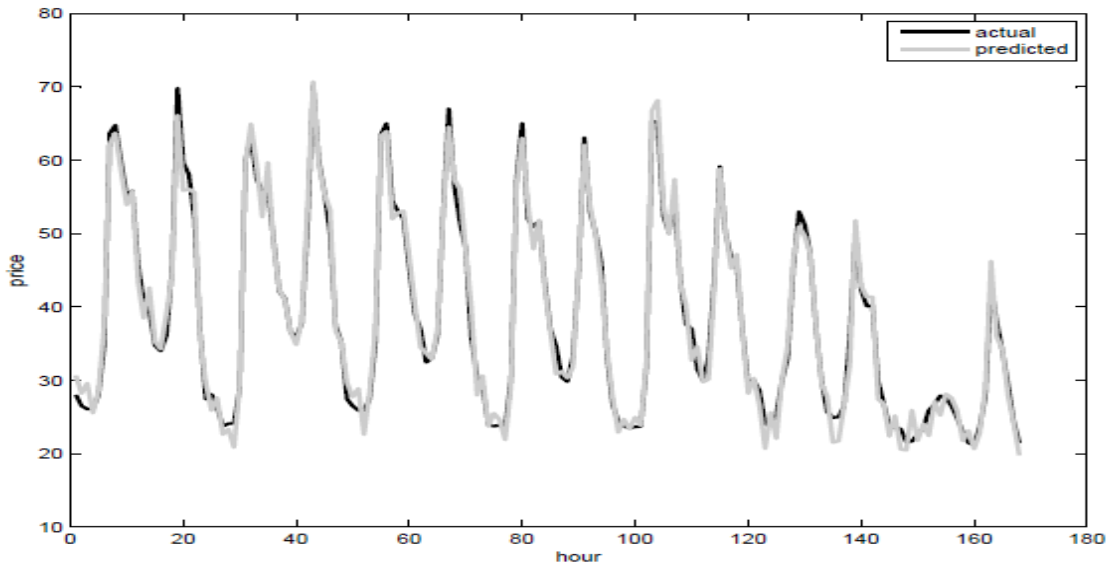


Fig.8.18 Predicted and Actual Prices for PJM Market for Week 1 for RNNFHN

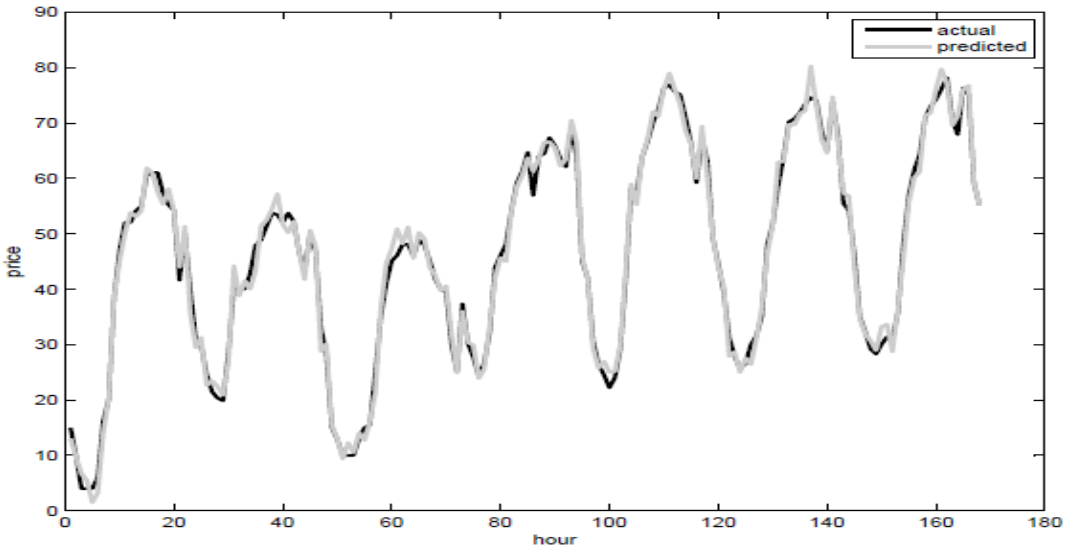


Fig. 8.19.Predicted and Actual Prices for PJM Market for Week 2 for RNNFHN

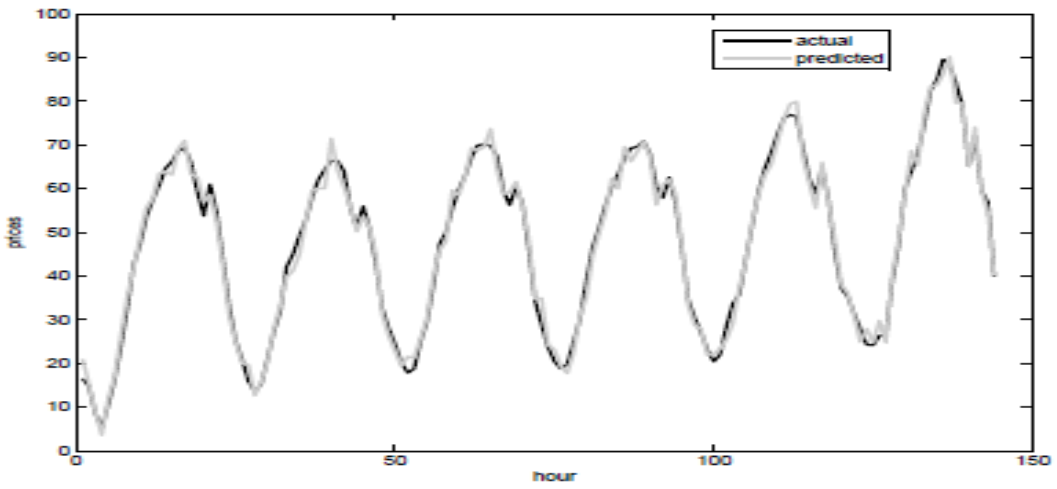


Fig. 8.20 Predicted and Actual Prices for PJM Market for Week 3 for RNNFHN

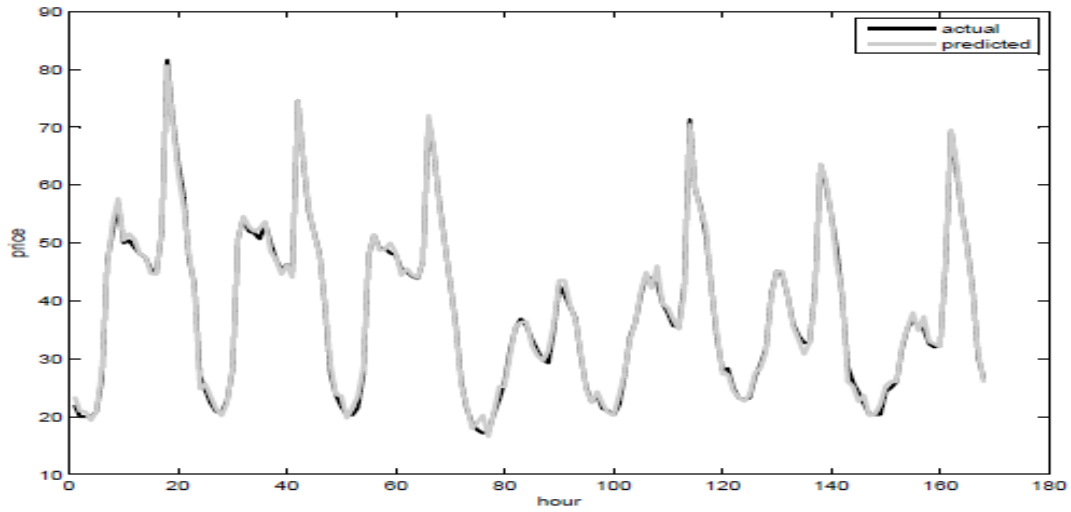


Fig. 8.21 Predicted and Actual Prices for PJM Market for Week 4 for RNNFHN

In the above four figures, it can be observed that the testing dataset for PJM market consists mainly of chaotic oscillations with almost no presence of spikes. The predicted and actual curve comparison confirms that RNNFHN works well in low dimensional chaotic

regions however results are inferior compared to benchmarks or PGRNN model.

8.4.3 Results for Victoria market

The implementation results for PGRNN for Victoria market are shown from Fig. 8.22-8.24 where actual and predicted curves are presented.

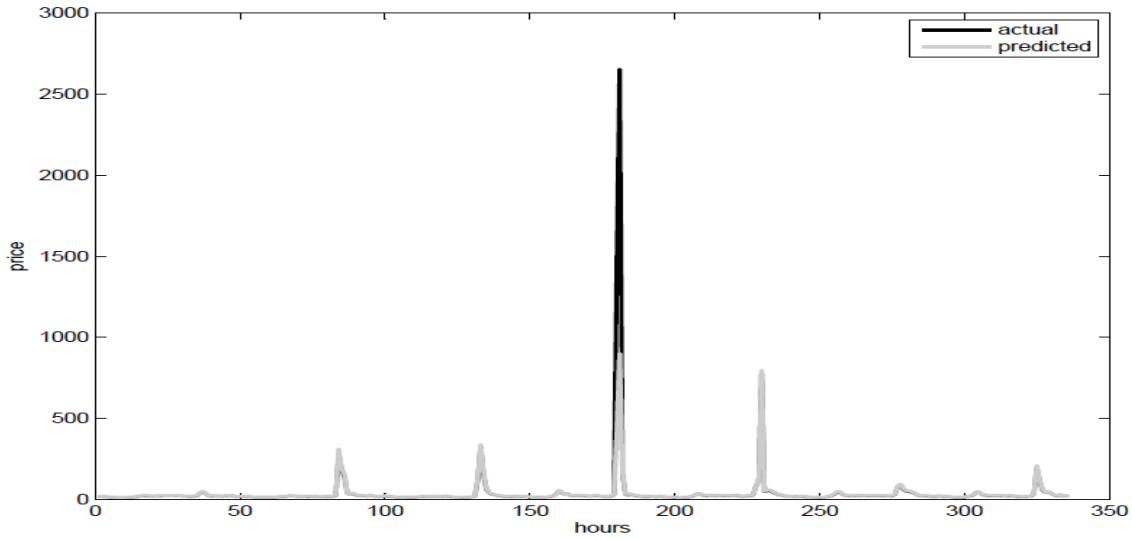


Fig. 8.22.Predicted and Actual Prices for Victoria Market for Week 1 for RNNFHN

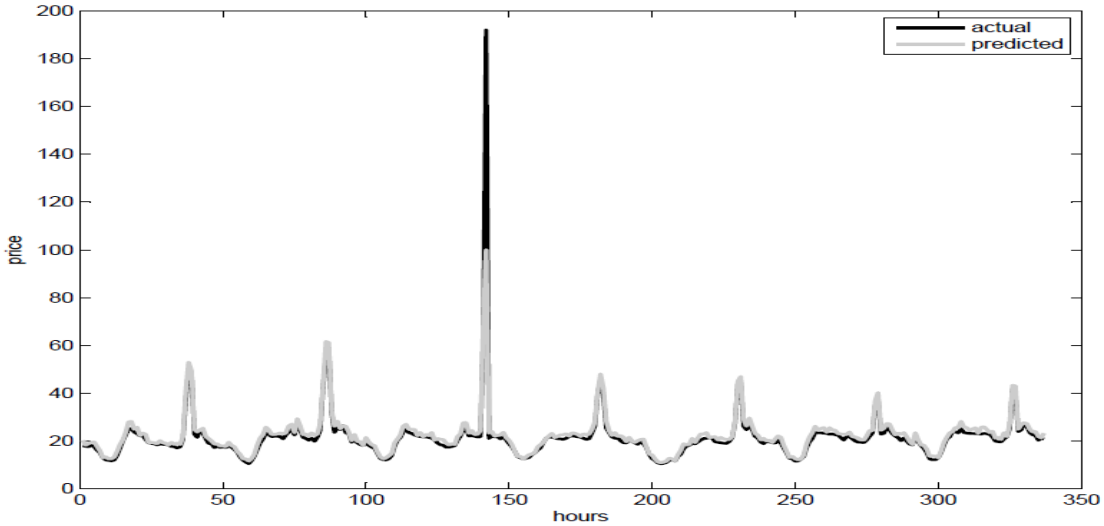


Fig. 8.23.Predicted and Actual Prices for Victoria Market for Week 2 for RNNFHN

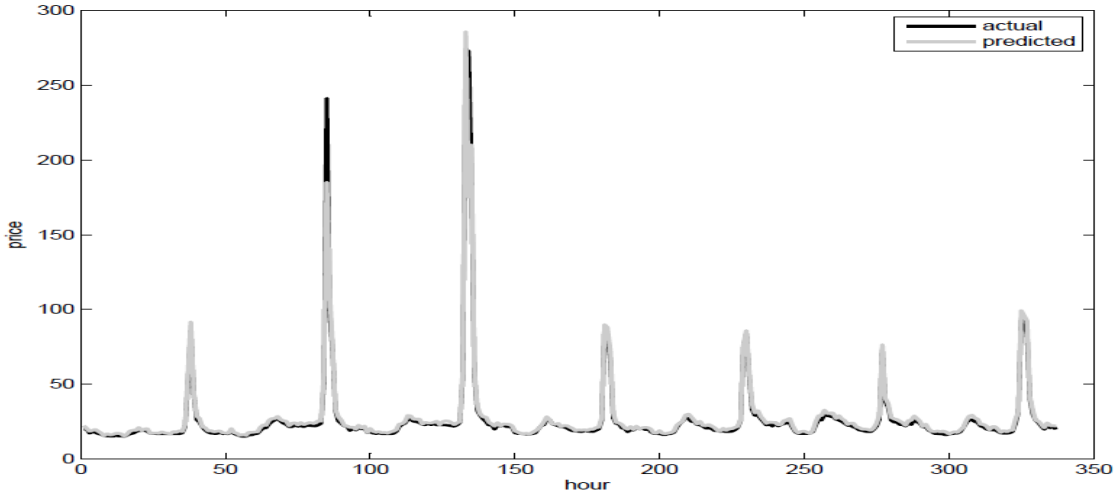


Fig. 8.24.Predicted and Actual Prices for Victoria Market for Week 3 for RNNFHN

In the above three figures, it can be observed that the testing dataset for Victoria market consists of low volatile chaos during week 1 and 2 while highly volatile chaotic oscillations in week 3 with intermittent presence of single spikes. The developed model

performs with considerable accuracy in both chaotic and spiking regions.

8.4.4 Results for NSW market

The implementation results for PGRNN for NSW market are shown from Fig. 8.25-8.28 where actual and predicted curves are presented.

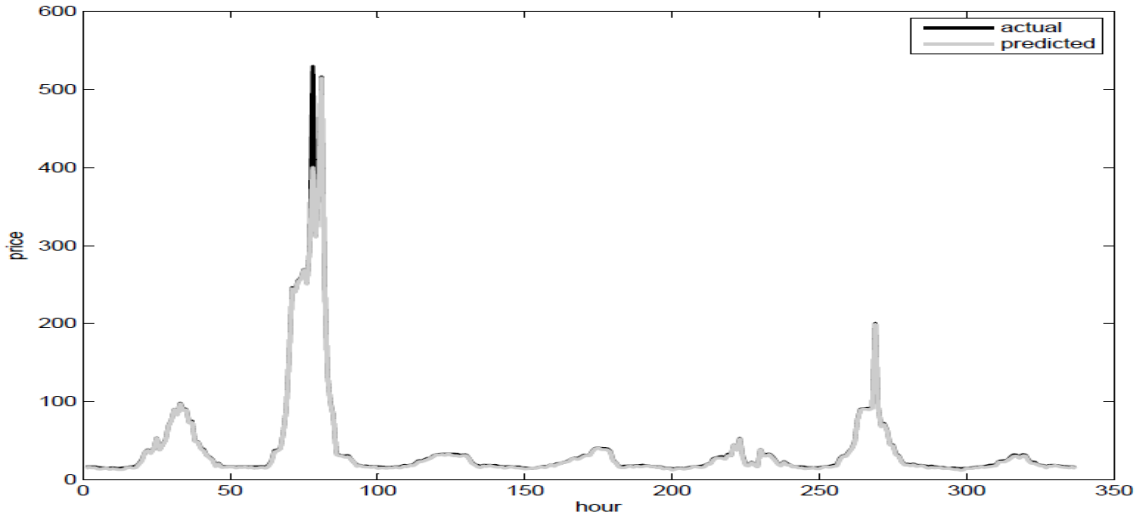


Fig. 8.25.Predicted and Actual Prices for NSW for Week 1 for RNNFHN

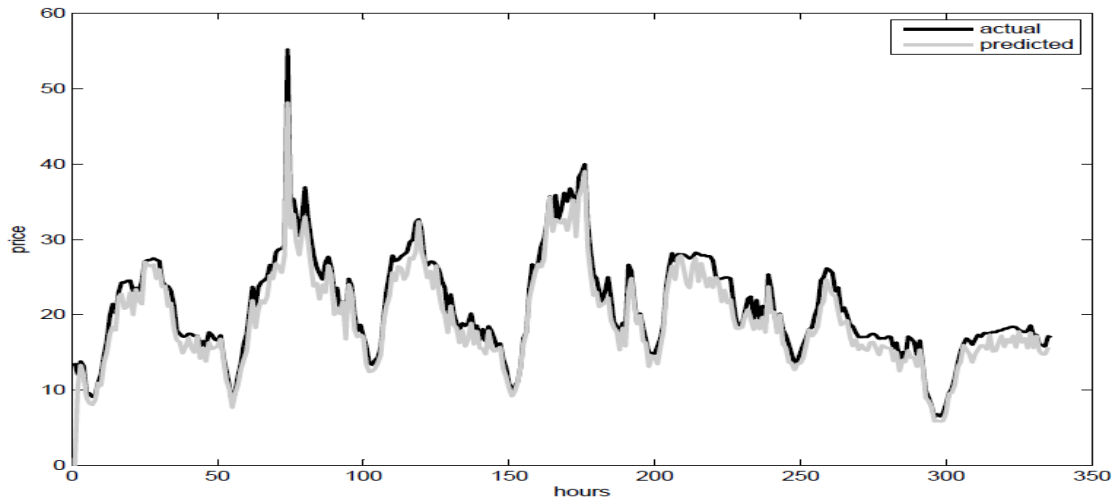


Fig. 8.26.Predicted and Actual Prices for NSW for Week 2 for RNNFHN

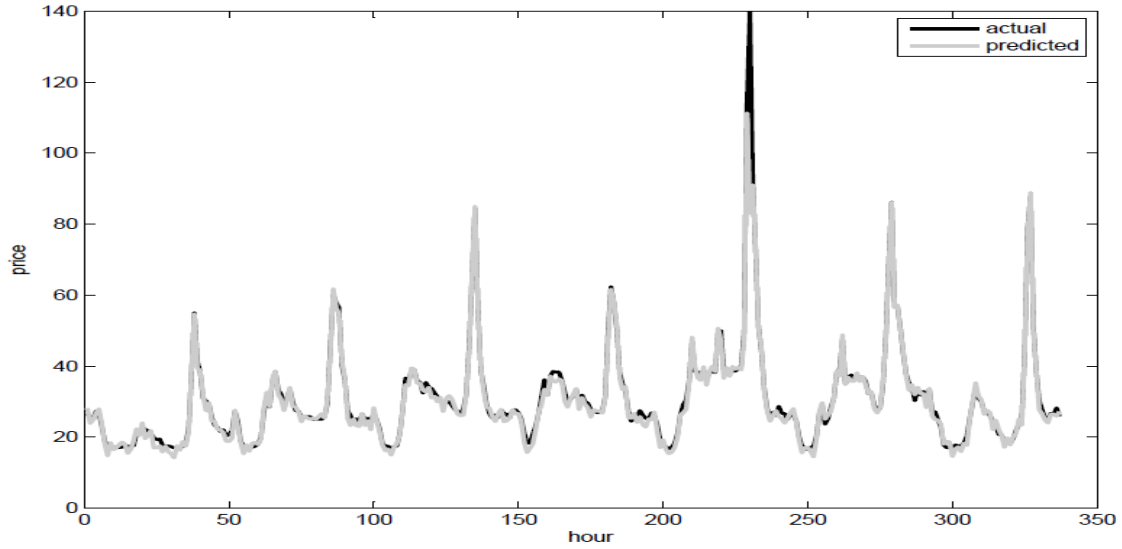


Fig. 8.27.Predicted and Actual Prices for NSW for Week 3 for RNNFHN

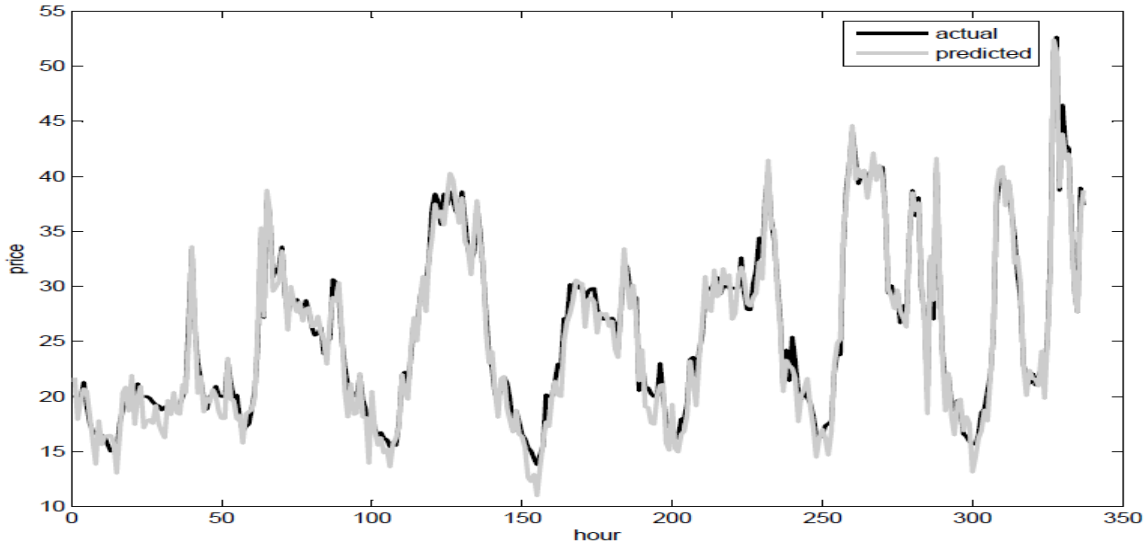


Fig. 8.28.Predicted and Actual Prices for NSW for Week 4 for RNNFHN

In the above four figures, it can be observed that the testing dataset for NSW market consists of low volatile chaos during week 1, 2 and 3 with intermittent spikes while highly volatile chaotic oscillations in week 4. However the unique feature of week 1 and week 4 is the nature of spikes which is in the form of low frequency bursting. The developed model performs with considerable accuracy in chaotic regions and single spiking regions, while the accuracy in bursting spiking region is much better compared to PGRNN.

8.5 MSRNN Implementation Results

The proposed model was implemented for forecasting of day-ahead prices in testing data. Time series data from each market was divided into training and testing sets. The corresponding performance curves, which show the comparison of predicted and actual results, are shown in Fig 8.29-8.42.

Table 8.6
Comparison of Proposed MSRNN and Benchmark Methods on Worldwide Deregulated Markets (Error is in **MAPE** for Ontario, NSW and NEMMCO and **WMAPE** for PJM)

Market	Season	Benchmark (%)	Proposed MSRNN
Ontario(week 1-2)	Spring	17.35 [161]	**12.83**
Ontario(week 3-4)	Summer	18.266 [161]	**12.9**
Ontario(week 5-6)	Winter	17.626 [161]	**12.42**
NEMMCO Victoria (Week1)	Summer	22.1 [162]	**10.4**
NEMMCO Victoria (Week2)	Summer	23.1 [162]	**10.96**
NEMMCO Victoria (Week3)	Winter	32.5 [162]	**13.86**
NEMMCO NSW(Week1)	Winter	15.57 [128]	**10.65**
NEMMCO NSW(Week2)	Fall	13.03 [128]	**9.78**
NEMMCO NSW(Week3)	Summer	13.84 [128]	**9.33**
NEMMCO NSW(Week4)	Spring	9.89 [128]	**8.60**
PJM (week 1)	Spring	**6.392 [127]**	6.42
PJM (week 2)	Fall	5.976 [127]	**5.01**
PJM (week 3)	Summer	5.974 [127]	**5.83**
PJM (week 4)	Winter	6.648 [127]	6.81

As observed in table 8.6, the proposed model exhibitS interesting results on tested markets. The RNN model trained using learning algorithm employing geometrical singular perturbation and fixed weight initialization based learning of fast subsystem gives best results for each tested deregulated markets. It can be observed that the accuracy obtained using the proposed model is better than benchmark models in most of the markets except

PJM in season of winter. However in case of Ontario market the results are particularly high, while the performance accuracy is quite good in Victoria markets. In NSW market the results are good in particular regions of time series.

It can be deduced that the proposed model accurately models the time series and the 24 hours ahead market can be forecasted using this approach. The corresponding predicted curves for all marktes are shown in figures 8.29 – 8.42 below.

8.5.1 Results for Ontario Market

The implementation results for MSRNN for Ontario market are shown from Fig. 8.29-8.31 where actual and predicted curves are presented.

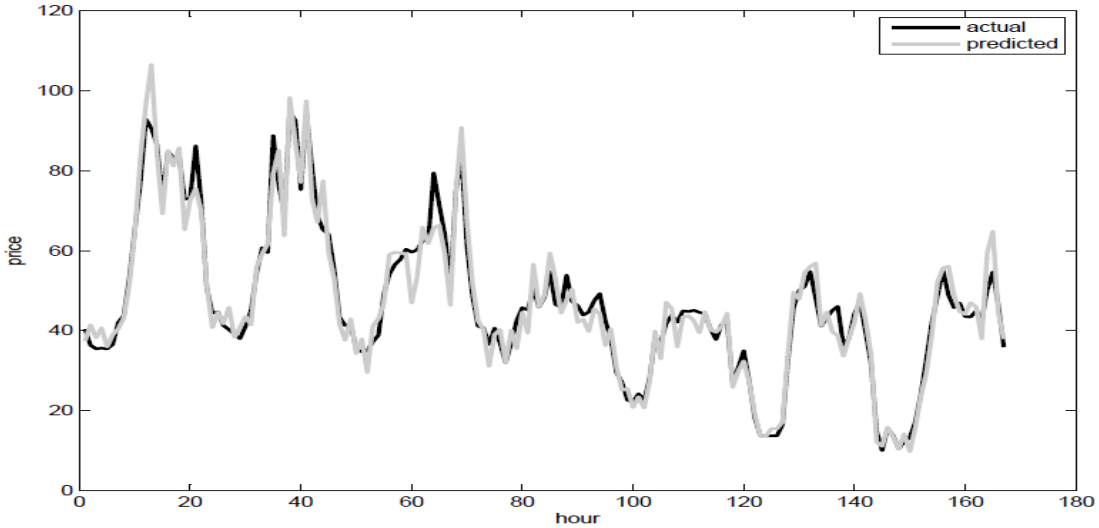


Fig.8.29.Predicted and Actual Prices for Ontario for Week1 for MSRNN

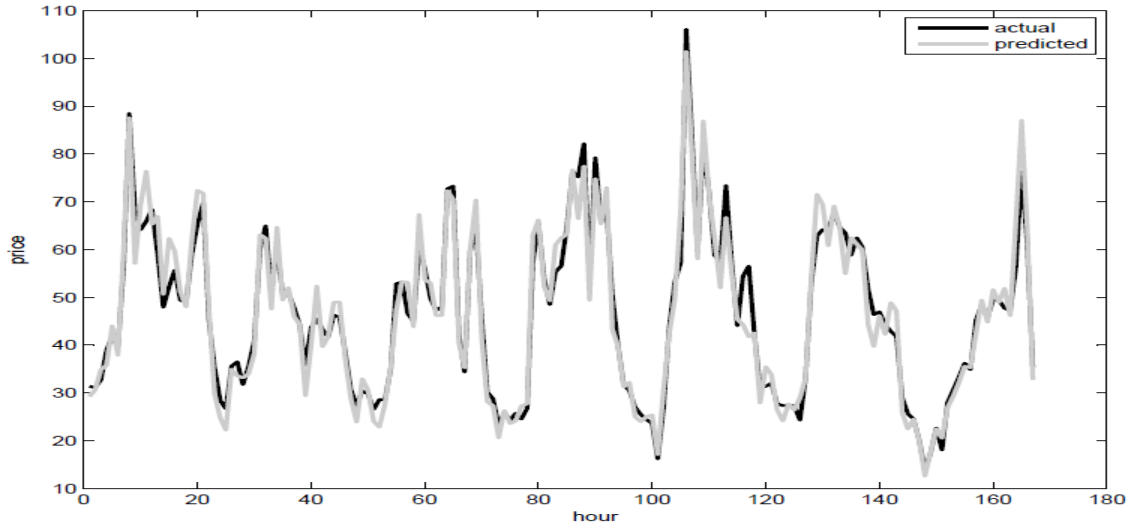


Fig.8.30. Predicted and Actual Prices for Ontario for Week2 for MSRNN

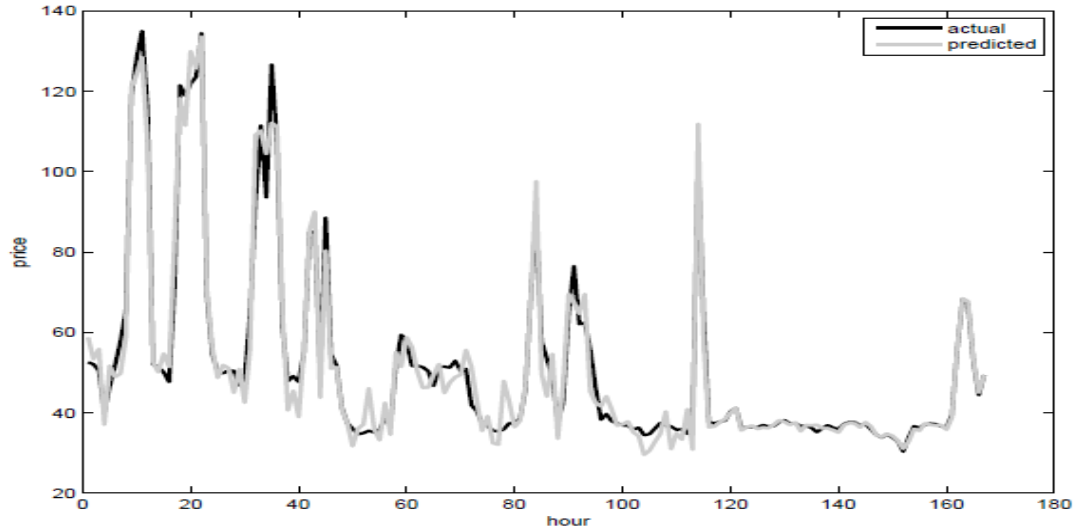


Fig.8.31. Predicted and Actual Prices for Ontario for Week3 for MSRNN

In the above three figures, it can be observed that the testing dataset for Ontario market consists of highly volatile chaotic oscillations with intermittent presence of single spikes. The developed model performs well in both spiking region and chaotic region with high accuracy.

8.5.2 Results for PJM market

The implementation results for PGRNN for PJM market are shown from Fig. 8.2-8.5 where actual and predicted curves are presented.

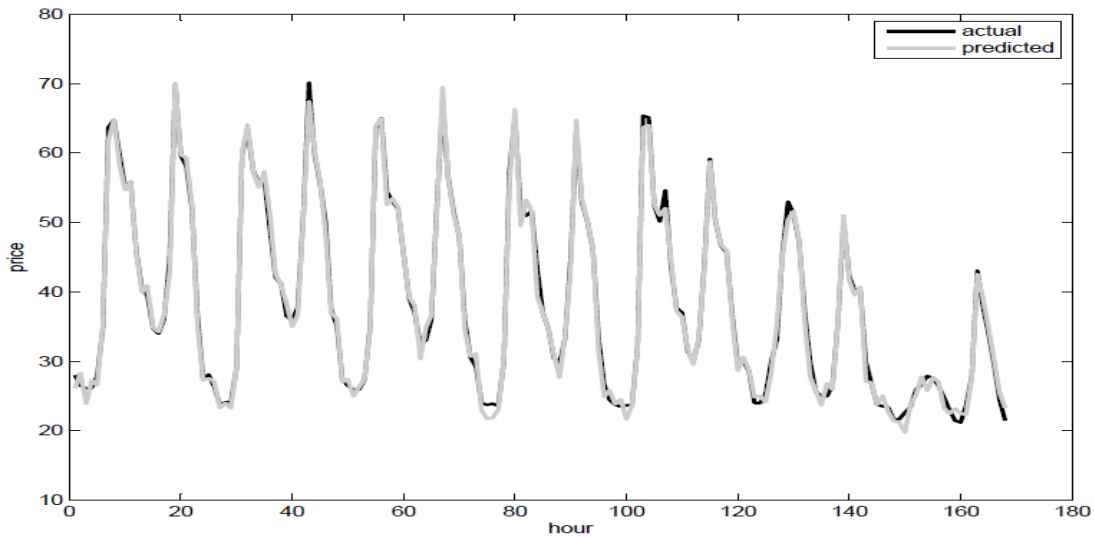


Fig.8.32. Predicted and Actual Prices for PJM for Week 1 for MSRNN

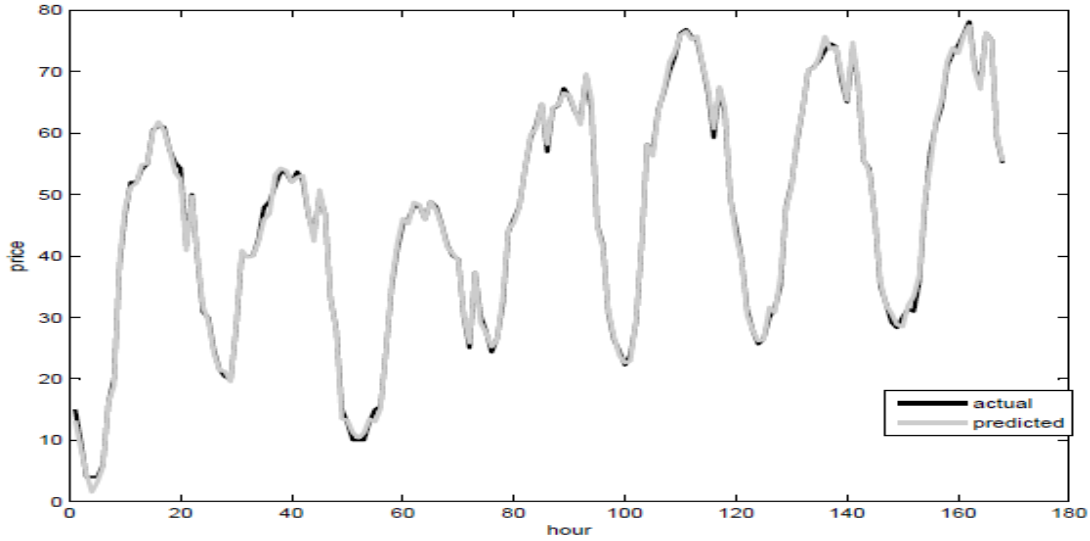


Fig.8.33. Predicted and Actual Prices for PJM for Week 2 for MSRNN

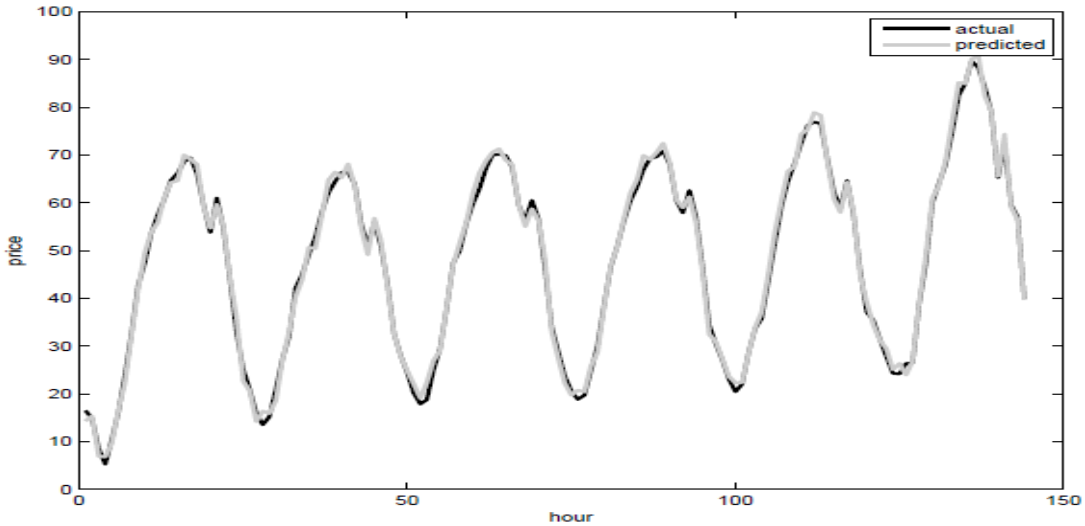


Fig.8.34 Predicted and Actual Prices for PJM for Week 2 for MSRNN

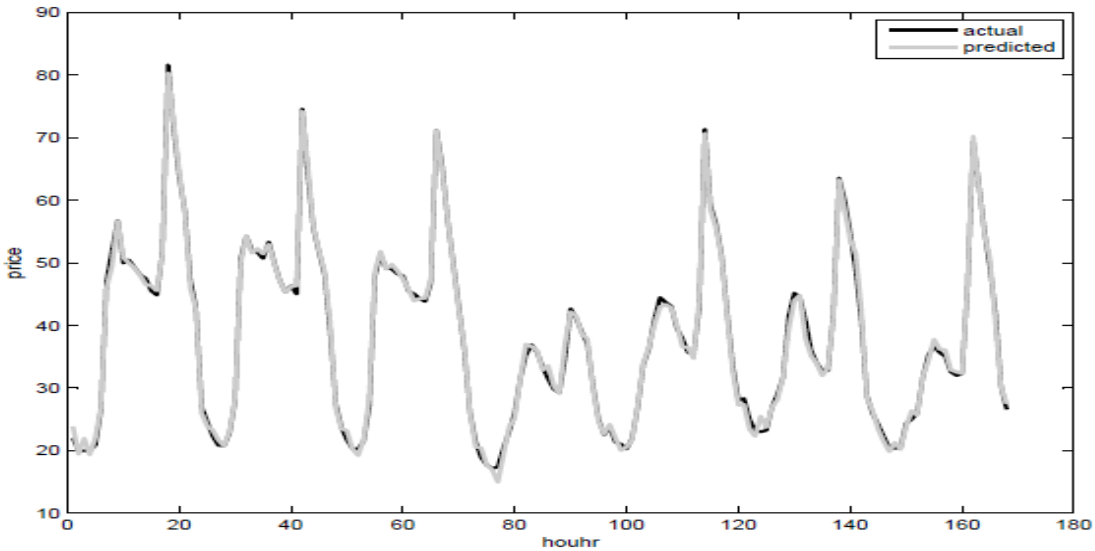


Fig.8.35. Predicted and Actual Prices for PJM for Week 3 for MSRNN

In the above four figures, it can be observed that the testing dataset for PJM market consists mainly of chaotic oscillations with almost no presence of spikes. The predicted and actual curve comparison confirms that RNNFHN works well in low dimensional chaotic regions however results are inferior compared to benchmarks or PGRNN model.

8.5.3 Results for Victoria Market

The implementation results for PGRNN for Victoria market are shown from Fig. 8.36-8.39 where actual and predicted curves are presented.

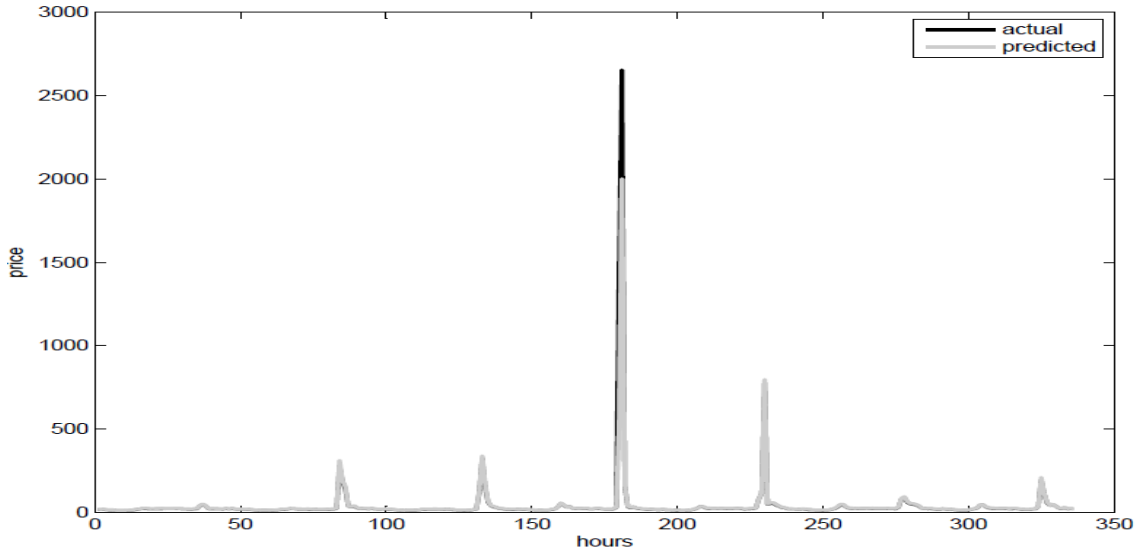


Fig.8.36. Predicted and Actual Prices for Victoria for Week 1 for MSRNN

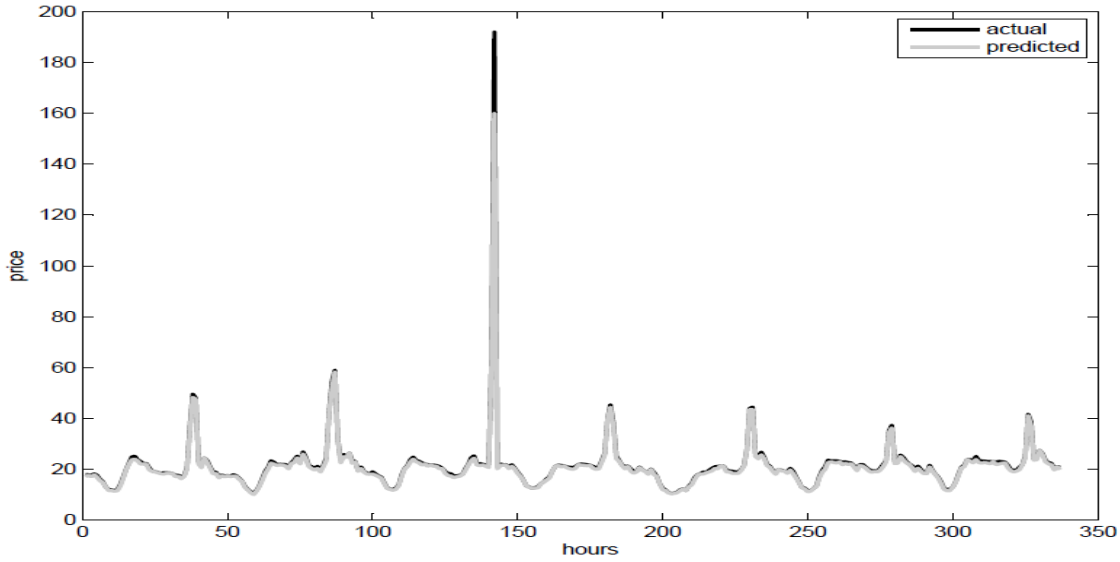


Fig.8.37. Predicted and Actual Prices for Victoria for Week 2 for MSRNN

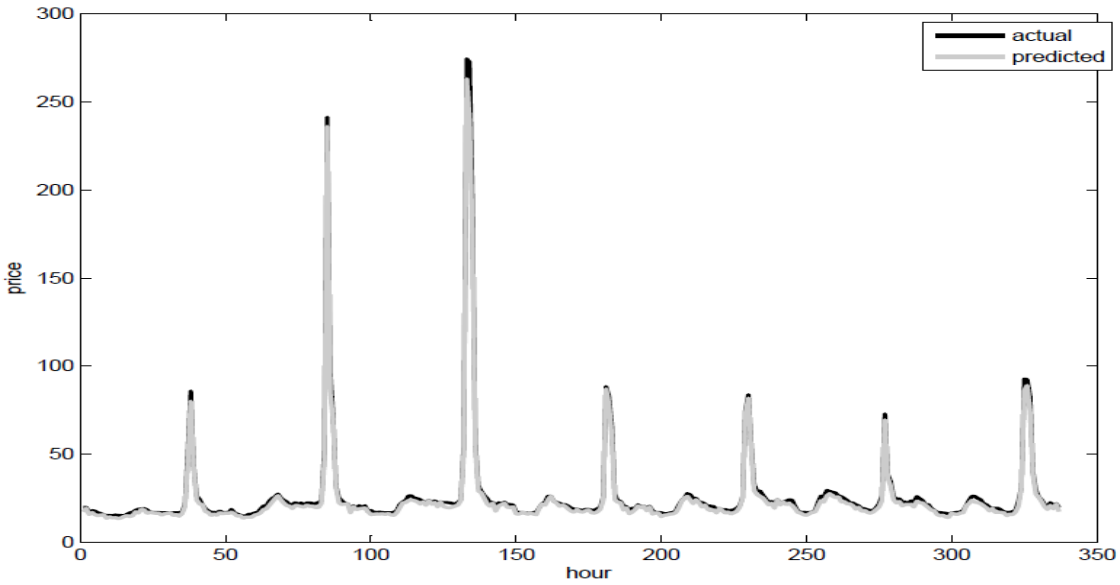


Fig.8.38. Predicted and Actual Prices for Victoria for Week 3 for MSRNN

In the above three figures, it can be observed that the testing dataset for Victoria market consists of low volatile chaos during week 1 and 2 while highly volatile chaotic oscillations in week 3 with intermittent presence of single spikes. The developed model performs with considerable accuracy in both chaotic and spiking regions.

8.5.4 Results for NSW market

The implementation results for PGRNN for NSW market are shown from Fig. 8.12-8.15 where actual and predicted curves are presented.

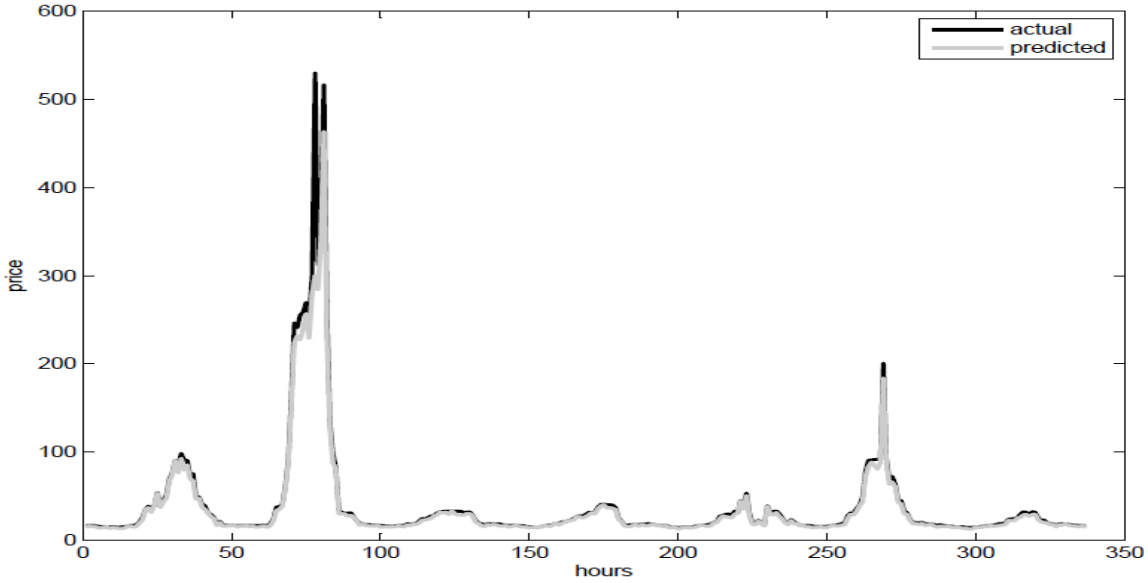


Fig.8.39. Predicted and Actual Prices for NSW for Week 1 for MSRNN

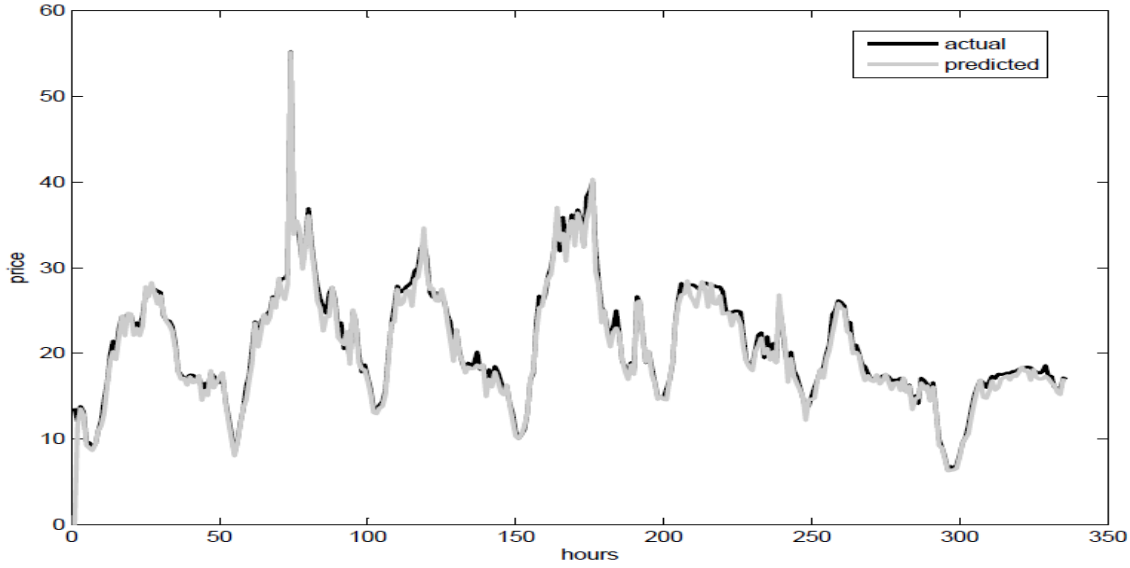


Fig.8.40. Predicted and Actual Prices for Victoria for Week 2 for MSRNN

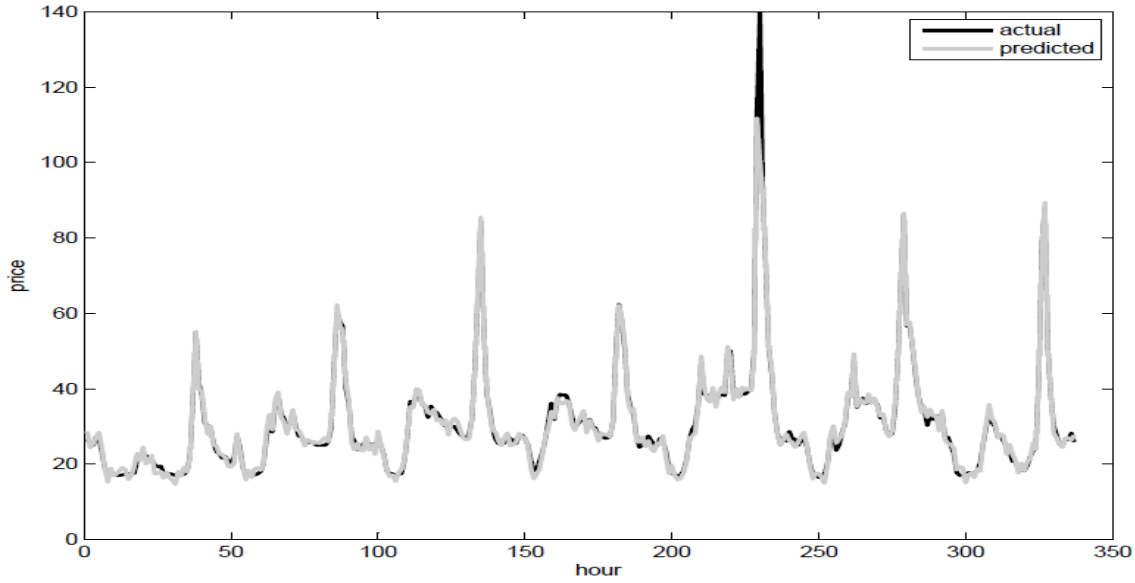


Fig.8.41. Predicted and Actual Prices for NSW for Week 3 for MSRNN

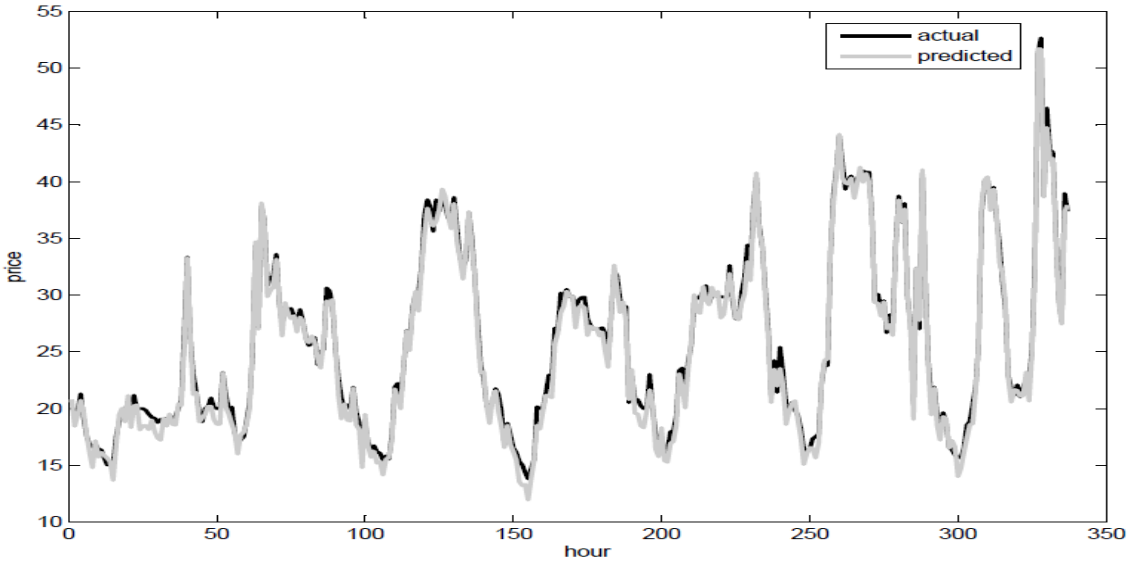


Fig.8.42. Predicted and Actual Prices for NSW for Week 4 for MSRNN

In the above four figures, it can be observed that the testing dataset for NSW market consists of low volatile chaos during week 1, 2 and 3 with intermittent spikes while highly volatile chaotic oscillations in week 4. However the unique feature of week 1 and week 4 is the nature of spikes which is in the form of low frequency bursting. The developed model performs with considerable accuracy in chaotic regions and single spiking regions, while the accuracy in bursting spiking region is much better compared to PGRNN.

8.6 Comparison of Performance of Three Proposed Models

The results from all three proposed approaches are compared in table 8.7 below.

Table 8.7
Comparison of the Proposed Approaches for Worldwide Deregulated Markets in Terms of Error Indices (Error is in **MAPE** for Ontario, NSW and NEMMCO and **WMAPE** for PJM)

Market	**Proposed RNN (approach 1) (%)**	**Proposed RNN-FHN (approach 2) (%)**	**Proposed MSRNN (approach 3) (%)**
Ontario(week 1-2)	13.43	**12.65**	12.83
Ontario(week 3-4)	12.96	**12.71**	12.9
Ontario(week 5-6)	12.75	**12.04**	12.42
NEMMCO Victoria (Week1)	12.61	13.5	**10.4**
NEMMCO Victoria (Week2)	12.99	14.02	**10.96**
NEMMCO Victoria (Week3)	14.17	15.66	**13.86**
NEMMCO NSW(Week1)	**10.19**	10.39	10.65
NEMMCO	**9.03**	11.28	9.78

NSW(Week2)			
NEMMCO NSW(Week3)	9.54	8.93	**9.33**
NEMMCO NSW(Week4)	**8.34**	9.53	8.60
PJM (week 1)	**5.29**	5.95	6.42
PJM (week 2)	5.34	5.99	**5.01**
PJM (week 3)	5.9	6.71	**5.83**
PJM (week 4)	**6.49**	6.65	6.81

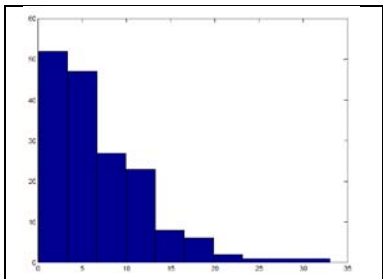
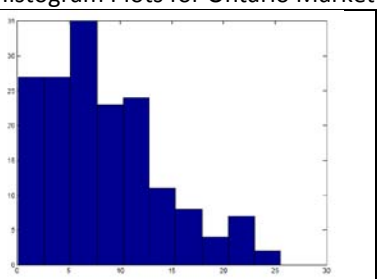
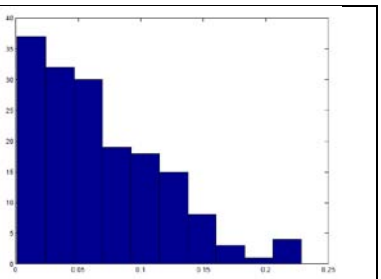
The above table exhibits some interesting results. It can be observed that RNN model trained using approach 1 and approach 3 perform well in most of the markets. However in Ontario market where the spiking behavior is a combination of isolated spikes and burst of spikes, the approach 2 based on RNN-FHN outperforms other two approaches, while approach 3, which is capable of predicting spikes, outperforms approach 1. In Victoria market, where the spiking transitions occur only in form of isolated spikes at irregular intervals, approach 3 based on MSRNN performs very well and outperforms other two approaches. In NSW market, where the inherent dynamics is complex chaos, the approach based on state feedback based RNN trained using invariant dynamics performs the best. Finally in PJM market, where the dynamics are relatively less complex compared to other markets, approach 1 and approach 3 perform well compared to approach 2. The evidence of above analysis can be seen in error histogram analysis in next section.

8.7 Error Histogram Analysis

The results shown above in terms of MAPE and WMAPE represent an important metric based evaluation, however they do not give a clear picture of error distribution over the test period. This distribution is important for decision makers in competitive markets, especially in case of spiky time series. In spiking scenarios, the high spike regions represent very high risk regions where high error in prediction can result in loss of millions of worth of utilities. Hence it is desired to have a lower error in spiking regions. These scenarios can be analyzed through error histogram plots which absolute error distribution along the time series. Hence an important desirable feature of error histogram is symmetric bar heights over moderately lower error regions with no bars in very high error range. This scenario is highly favored compared to tall bars of very low error regions and short bars over corresponding to very high error regions.

The error histogram plots are shown in Table 8.8 – 8.10 where the first, second and third column represents results corresponding to PGRNN, FHNRNN and MSRNN.

Table 8.8
Error Histogram Plots for Ontario Market

		
PGRNN Week 1	RNNFHN Week 1	MSRNN Week 1

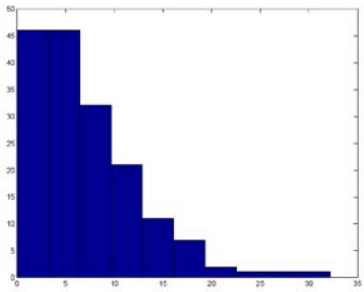
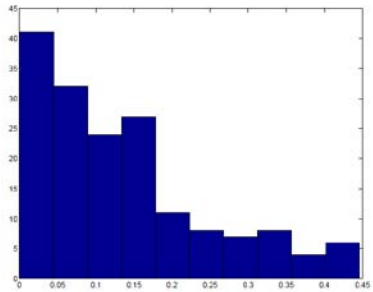
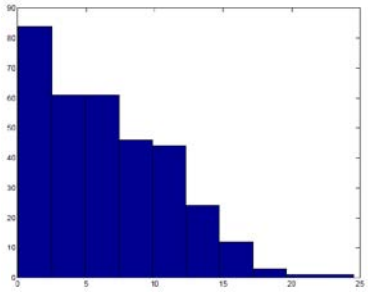
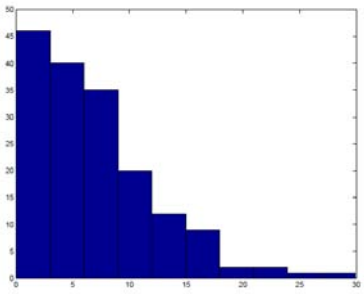
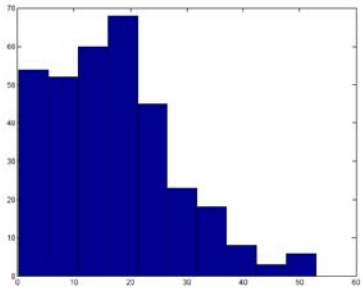
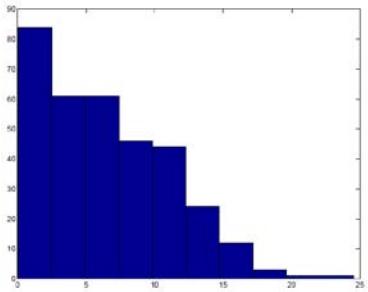
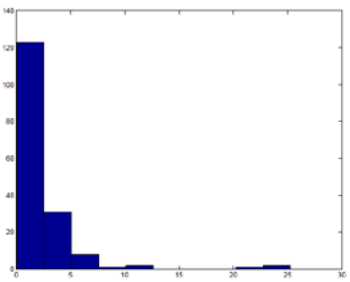
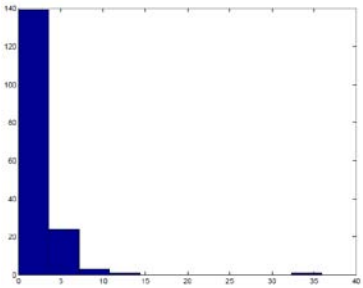
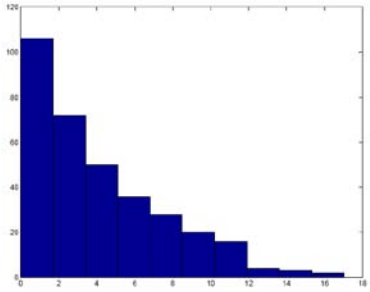
		
PGRNN Week 2	RNNFHN Week 2	MSRNN Week 2
		
PGRNN Week 3	RNNFHN Week 3	MSRNN Week 3

Table 8.9
Error Histogram Plots for Victoria Market

		
PGRNN Week 2	RNNFHN Week 2	MSRNN Week 2

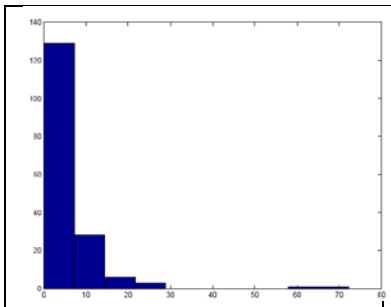
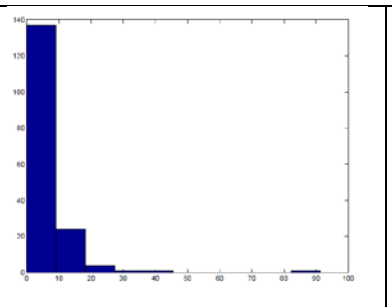
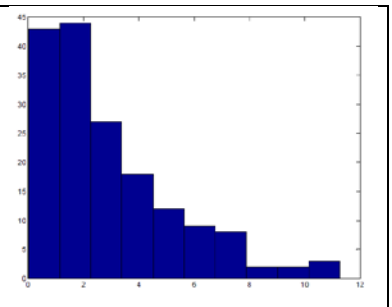
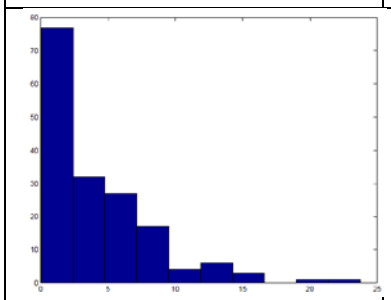
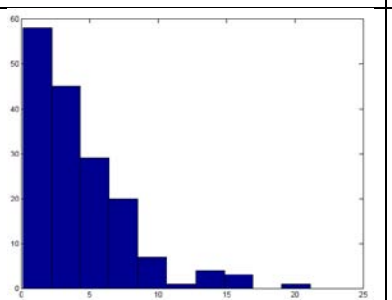
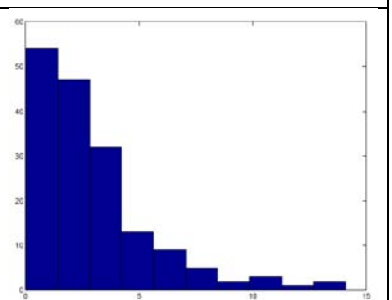
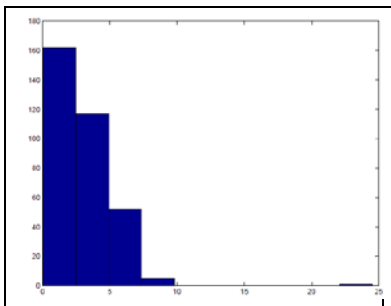
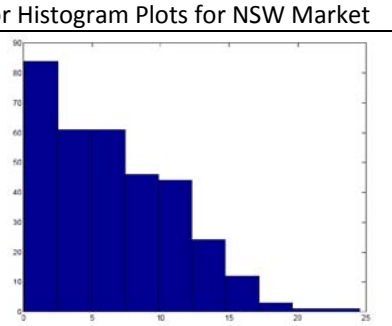
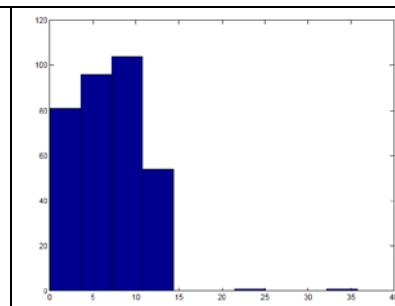
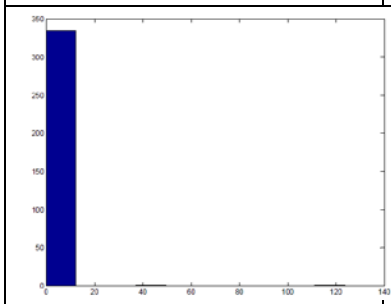
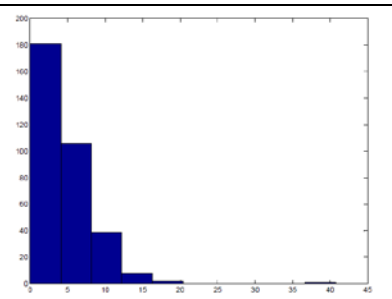
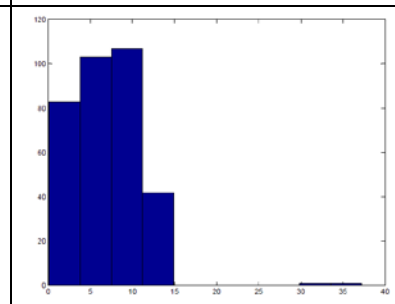
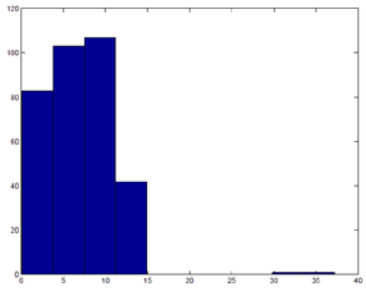
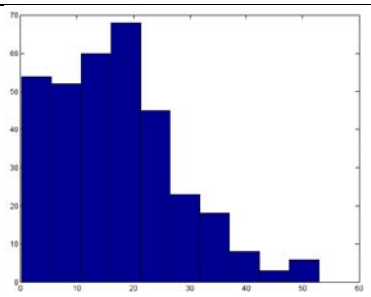
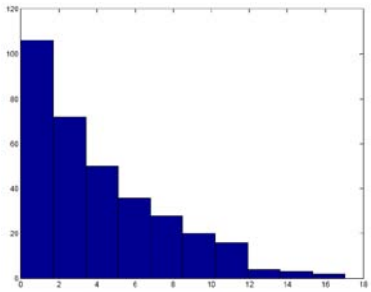
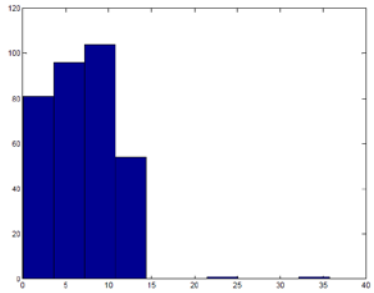
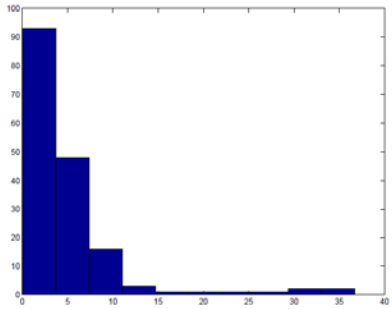
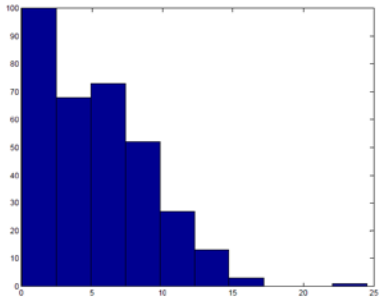
		
PGRNN Week 3	RNNFHN Week 3	MSRNN Week 3
		
PGRNN Week 4	RNNFHN Week 4	MSRNN Week 4

Table 8.10
Error Histogram Plots for NSW Market

		
PGRNN Week 1	RNNFHN Week 1	MSRNN Week 1
		

PGRNN Week 2	RNNFHN Week 2	MSRNN Week 2
		
PGRNN Week 3	RNNFHN Week 3	MSRNN Week 3
		
PGRNN Week 4	RNNFHN Week 4	MSRNN Week 4

The histogram plots allow examination of error characteristics of each model. It can be seen that best error characteristics are exhibited by PGRNN for Ontario market, MSRNN for Victoria market, and RNNFHN for NSW market. The explanation for this behaviour is similar to discussion above, i.e. the difference in models performance in chaotic and spiking regimes.

8.8 Discussion

In this chapter implementation results for three novel approach is discussed.

In first approach the invariant dynamics of time series are incorporated during the learning process (in particular fixed point dynamics) and enforced during each step of

learning. This approach was proposed in attempt to achieve a dynamic RNN model which learns the inherent complex nonlinearities of the system which makes it a closer approximation of actual nonlinear system. This approach can also be seen as using heuristic information to achieve global optimum solution during time series based RNN learning. The three steps of learning process include intelligent weight initialization using fixed point dynamics, enforcing fixed point conditions during each learning step and enforcing local jacobian during the course of time series. The trained model outperforms the recent benchmark approaches in most of the deregulated markets. The results of all three steps of incorporating invariant dynamics are shown individually in table 8.3 . It can be observed that the weight initialization step is the most contributing factor towards improvement in results using RNN. The projection gradient descent contributes in improving the result in case of hyperchaotic systems such as Ontario time series, which implies that the Ontario time series system has strong but stationary chaotic behaviour in most of the region of phase space. The contribution of learning of local jacobian during the training process is not very significant although this step helps in learning the local variations in time series.

For second approach, the performance curves and the comparison table indicates the superiority of proposed model over other forecasting approaches proposed in past, over specified testing dataset. Moreover closer prediction in spiking region strengthens its applicability as forecasting tool in high risk electricity markets. The main contributions of this thesis and the key advantages of the proposed model can be summarized as follows.

- Improved prediction performance achieved through development of a heuristic based hybrid model combining merits of RNN models and excitable properties of FHN coupled system of equations with better forecasting.

- Satisfactory tracking and forecast in the spiking region which helps in good statistical results, thereby reducing the financial risk in this region of uncertainty.
- Reasonably accurate performance over worldwide deregulated markets from different regions of the world over different seasons and time of the year, verifying the adaptability and applicability of the proposed model.
- Satisfactory unconditional coverage provided by obtained prediction intervals upon interval forecasting. Prediction intervals obtained are narrow in stable regime and wider in spiky volatile regions.
- The developed model avoids preprocessing of time series data, thereby avoiding possible loss of data due to filtering. In particular, avoiding the decomposition of time series signal averts the risk of losing high frequency components.
- The developed model handles the problem of non-stationarity, which is why it doesn't necessitate limiting the size of training data.

While the proposed model is able to achieve better results compared to traditional models on this time series, the MAPE of the forecasts can be further improved. The absence of information about the structure of the network required is a small disadvantage with this approach, but the key advantage of using this model is the ability to achieve fair predictions in the spiking region, which will help the bidders and strategy makers a great deal at the critical time.

In third approach, the forecasting performance of this model in the spiking and nonspiking region was found to be excellent. Success of this MSRNN model presents exciting opportunities for future work in this time series system. Further work will include extensive investigation of slow-fast scale analysis of electricity price time series and include the results as heuristics in modelling of MSRNN.

8.9 Limitations of Developed Models

In this chapter it was observed that the three developed models perform better compared to benchmark approaches. However it can be seen that the volatility structure of all deregulated markets and the three models underperform in few cases. The limitations of all proposed model are summarized and discussed below

8.9.1 Limitation of PGRNN

- From the results it can be observed that the model doesn't predict and model the spikes accurately.
- One major limitation of this model is modelling complexity and dependence of model on accuracy of extracting fixed point of time series.
- Another limitation of this model is that it is not adaptable in case of occurrence of non-stationarity.

8.9.2 Limitation of RNNFHN

- From results it can be observed that the model doesn't perform well in chaotic region of time series particularly.
- Due to unique hybrid nature of this model, it is difficult to prove the stability of model.

8.9.3 Limitation of MSRNN

- From the results it can be observed that this model perform well in spiking as well as chaotic regime, however the model underperforms in cases when spiking behaviour deviates from single spikes to bursting spikes.

- Another limitation of this approach is separate training of slow subsystem and fast subsystem of MSRNN. This results in slower learning process which renders the system inapplicable in cases where real time learning is required.

Chapter 9

Conclusion and Future Work

In this thesis, a systematic approach based on nonlinear dynamics has been adopted towards successful modelling of spiking time series in deregulated electricity markets using Recurrent Neural Networks. The first step involved analyzing time series using nonlinear systems theory and observing the invariants measures constituting time series which is presented in chapter 4. Embedding dimension is the basic invariant measure which is crucial for reconstruction of phase space of time series. Furthermore, the lyapunov exponent of time series is calculated to measure the degree of chaos in the system. However due to presence of possible intermittent variation in dynamics, lyapunov exponent is not a reliable index, which is why Finite Size Lyapunov Exponent (FSLE) and Scale Dependent Lyapunov Exponent (SDLE) are calculated to analyze the transient dynamics of time series along with global behavior. Observing FSLE allows in detecting the possible presence of loss of hyperbolicity in the system. Moreover it is proven that spiking mechanism doesn't occur due to loss of local hypervbolicity, or UDV. SDLE analysis allows establishing the fact that electricity price is not a noisy time series, on the other hand exhibits complex irregular behavior due presence of dynamics on multiple scales. Most importantly, fixed point location and the dynamics in neighbourhood are analyzed which are the crucial invariant of time series. It was observed that all time series are nonlinear systems with presence of saddle. Eigenvalues of these saddles determine the behavior of the system globally and locally. In chapter 5, RNN is developed which employ the fixed point dynamics during learning process. The proposed approach results in better learning of RNN by bring the RNN closer to exact nonlinear system underlying time series. The results indicated that although this model outperforms classically trained RNN and benchmark models, and perform quite

well in the interspike chaotic region, the results in spiking region still required improvement. In the next two chapters we pay particular attention on modelling spiking region of time series. From dynamical system perspective, spiking transition is approached from multiple scale behavior. We describe the spiking behavior as critical transition in a multiple scale system where the system dynamics bifurcate due to variation in “parameter” at slower scale.

In chapter 6, one of the dynamical systems exhibiting multiple scale dynamics is used to study spiking, the well known Fitz-Hugh Nagumo (FHN). A proposed hybridized model with recurrent neural network (RNN) is used for forecasting in electricity price time series. First the presence of multiple scale dynamics in the time series system is confirmed. The approach adopts extraction of dynamics of slow parameter system, variations in which affect the behavior of fast actual time series. Further a dynamic system with slow and fast scales, namely Fitz-Hugh Nagumo (FHN), is used and hybridized with recurrent neural networks. The property of the multiple scale equation system allows the mechanism of spiking in such regimes. In turn, the parameters and coupling variables of this excitable system are determined using an RNN based model. As a result the learned hybrid model would achieve a desired level of modeling accuracy. The developed hybrid model was tested in various markets worldwide over different seasons to test its forecasting ability, adaptability and robustness. Most volatile electricity markets, California, Australia, PJM, Spain and Ontario market in Canada were modeled using the proposed approach. Extensive comparative studies suggest that our approach yields favorite results in hour-ahead and day ahead price prediction and day-ahead market.

In chapter 7, for more accurate modeling of time series and the associated slow fast dynamics, a multiple scale neural network (MSNN) is developed. Slow fast systems deal with

slow manifold and fast manifold where the key dynamics of time series occur on fast invariant manifold while the dynamics occurring on slow manifold is responsible for intermittent critical transitions. The developed model is trained using singular perturbation theory for slow-fast systems combined with gradient descent algorithm. The results from all three proposed approaches are compared in table 9.1 below.

Table 9.1
Comparison of the Proposed Approaches for Worldwide Deregulated Markets in Terms of Error Indices (Error is in **MAPE** for Ontario, NSW and NEMMCO and **WMAPE** for PJM)

Market	**PGRNN (approach 1) (%)**	**RNNFHN (approach 2) (%)**	**MSRNN (approach 3) (%)**
Ontario(week 1)	13.43	**12.65**	12.83
Ontario(week 2)	12.96	**12.71**	12.9
Ontario(week 3)	12.75	**12.04**	12.42
NEMMCO Victoria (Week1)	12.61	13.5	**10.4**
NEMMCO Victoria (Week2)	12.99	14.02	**10.96**
NEMMCO Victoria (Week3)	14.17	15.66	**13.86**
NEMMCO NSW(Week1)	**10.19**	10.39	10.65
NEMMCO NSW(Week2)	**9.03**	11.28	9.78
NEMMCO NSW(Week3)	9.54	8.93	**9.33**

NEMMCO NSW(Week4)	**8.34**	9.53	8.60
PJM (week 1)	**5.29**	5.95	6.42
PJM (week 2)	5.34	5.99	**5.01**
PJM (week 3)	5.9	6.71	**5.83**
PJM (week 4)	**6.49**	6.65	6.81

The above table exhibits some interesting results. It can be observed that RNN model trained using approach 1 and approach 3 perform well in most of the markets. However in Ontario market where the spiking behavior is a combination of isolated spikes and burst of spikes, the approach 2 based on RNN-FHN outperforms other two approaches, while approach 3, which is capable of predicting spikes, outperforms approach 1. In Victoria market, where the spiking transitions occur only in form of isolated spikes at irregular intervals, approach 3 based on MSRNN performs very well and outperforms other two approaches. In NSW market, where the inherent dynamics is complex chaos, the approach based on s tate feedback based RNN trained using invariant dynamics performs the best. Finally in PJM market, where the dynamics are relatively less complex compared to other markets, approach 1 and approach 3 perform well compared to approach 2.

9.1 List of Achievements

In summary, the list of achievements in this thesis can be given as

- The presence of multiple scale dynamics in electricity price time series was confirmed in chapter 4.
- From SDLE analysis it was validated that noise does not play crucial part in

dynamics of time series.

- A novel learning algorithm for training RNN was developed in chapter 5, which incorporates invariant dynamics of time series in learning process. The developed model achieved high accuracy in interspike chaotic regime.
- A multiple scale neural network was developed in chapter 7 and trained using singular perturbation theory and the novel RNN learning algorithm developed in chapter 5. The developed model achieved high accuracy in both spiking and interspike chaotic regime.

9.2 Future Work

The proposed approaches were developed to learn the complex inherent dynamics of electricity price time series in deregulated markets, along with the spiking behavior. The approaches perform better over worldwide markets compared to benchmarks. Interesting avenues of future research can be found in MSRNN and PGRNN. The limitations of these approaches are discussed in section 8.9. The main shortcoming of approach based on PGRNN is the inability to handle spikes successfully; however they perform quite well in highly volatile chaotic/hyperchaotic scenarios. The future work employing PGRNN will include application forecasting in case of other financial time series. Competitive and exchange rate markets are complex systems with large number of market players rendering highly volatile behavior. However the existing forecasting models do not perform satisfactorily to be employed with high confidence by market players. In order to implement PGRNN in these systems, it is of utmost importance to perform stability analysis of the developed learning algorithm. Hence the future work will include extensive stability analysis

of PGRNN learning algorithms.

While approach based on MSRNN performs very well in all markets with both spiking and chaotic regimes and outperform all benchmark approaches, they perform well in case of islolated spiking scenarios. It was observed that MSRNN doesn't capture the complex spiking patterns such as mixture of isolated spike and bursts, which are associated to complex homoclinic/heteroclinic scenarios. The future work will include extraction of these dynamics from time series. Investigation of these dynamics will include sophisticated analysis of variation in fixed point dynamics as the slow scale parameter variation takes place. This involves identification of exact slow manifold from time series and analysis of system using geometrical singular perturbation theory. This analysis and the obtained dynamic information can be incorporated during learning of multiple-scale RNN.

To summarize, the future work will broadly include.

- Application of MSRNN and PGRNN on other financial systems such as stock markets, given the excellent approximation capabilities of the model in the complex chaotic scenarios.
- Extensive stability analysis of PGRNN.
- Improving prediction performance of MSRNN by incorporating exact homoclinic/heteroclinic scenarios in the learning process.

REFERENCES

[1] , A. Ramos, and M. Rivier, "Electricity market modeling trends," *Energy Policy,* vol. 33, pp. 897-913, 2005.

[2] I. Vehviläinen and T. Pyykkönen, "Stochastic factor model for electricity spot price—the case of the Nordic market," *Energy Economics,* vol. 27, pp. 351-367, 2005.

[3] A. J. Conejo, J. Contreras, R. Espinola, and M. A. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market," *International Journal of Forecasting,* vol. 21, pp. 435-62, 2005.

[4] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola, "Forecasting next-day electricity prices by time series models," *IEEE Transactions on Power Systems,* vol. 17, pp. 342-8, 2002.

[5] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Transactions on Power Systems,* vol. 18, pp. 1014-20, 2003.

[6] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Transactions on Power Systems,* vol. 20, pp. 1035-42, 2005.

[7] M. Davison, C. L. Anderson, B. Marcus, and K. Anderson, "Development of a hybrid model for electrical power spot prices," *IEEE Transactions on Power Systems,* vol. 17, pp. 257-64, 2002.

[8] S. Stoft, , , *Power System Economics: Designing Markets for Electricity,*: Wiley/IEEE Press, May, 2002.

[9] C. R. Knittel and M. R. Roberts, "An Empirical Examination of Deregulated Electricity Prices," *SSRN eLibrary,* 2001.

[10] J. S. McMenamin. Statistical approaches to electricity price forecasting. http://capabilities.itron.com/downloads/pjmprices.

[11] T. B. A. Robinson, "- The Volatility of Prices in the English and Welsh Electricity Pool," *- Applied Economics. Vol. 34 (12). p 1487-95. August 2002.*

[12] T. A. Robinson, "Electricity pool prices: a case study in nonlinear time-series modelling," *Applied Economics,* vol. 32, pp. 527-532, Apr 2000.

[13] Á. Escribano Saez, J. I. Peña, and P. Villaplana "Modeling Electricity Prices: International Evidence," *SSRN eLibrary,* 2002.

[14] S. Deng, "Pricing electricity derivatives under alternative stochastic spot price models," presented at the 33rd Hawaii International Conference on System Sciences, 2000.

[15] A. S. De Vany and W. D. Walls, "Cointegration analysis of spot electricity prices: insights on transmission efficiency in the western US," *Energy Economics,* vol. 21, pp. 435-448, 1999.

[16] S. E. Bodily and M. Del Buono, "Risk and Reward at the Speed of Light: A New Electricity Price Model," *SSRN eLibrary,* 2002.

[17] D. Duffie, S. F. a. Gray, and P. H. Hoang, "Volatility in energy prices," in *Managing Energy Price Risk: The New Challenges and Solutions*. vol. 3rd V. Kaminski, Ed., ed London: Risk Books 2004.

[18] M. Ranjbar, S. Soleymani, N. Sadati, and A. M. Ranjbar, "Electricity Price Forecasting Using Artificial Neural Network," in *Power Electronics, Drives and Energy Systems, 2006. PEDES '06. International Conference on*, 2006, pp. 1-5.

[19] P. Mandal, T. Senjyu, N. Urasaki, and T. Funabashi, "A neural network based several-hour-ahead electric load forecasting using similar days approach," *International Journal of Electrical Power and Energy Systems,* vol. 28, pp. 367-373, 2006.

[20] A. I. Arciniegas and I. E. Arciniegas Rueda, "Forecasting short-term power prices in the Ontario Electricity Market (OEM) with a fuzzy logic based inference system," *Utilities Policy,* vol. 16, pp. 39-48, 2008.

[21] R. Weron, M. Bierbrauer, and S. Truck, "Modeling electricity prices: jump diffusion and regime switching," in *XVIII Max Born Symposium. Statistical Physics Outside Physics, 22-25 Sept. 2003*, Netherlands, 2004, pp. 39-48.

[22] A. Surana and G. Haller, "Ghost manifolds in slow–fast systems, with applications to unsteady fluid flow separation," *Physica D: Nonlinear Phenomena,* vol. 237, pp. 1507-1529, 2008.

[23] S. Haykin, *Neural Networks: A Comprehensive Foundation*: Prentice Hall, 1992.

[24] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, "Tipping elements in the Earth's climate system," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, pp. 1786-1793, Feb 2008.

[25] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter, and G. Vachtsevanos, "Epileptic seizures may begin hours in advance of clinical onset: A report of five patients," *Neuron,* vol. 30, pp. 51-64, Apr 2001.

[26] P. E. McSharry, L. A. Smith, and L. Tarassenko, "Prediction of epileptic seizures: are nonlinear methods relevant?," *Nature Medicine,* vol. 9, pp. 241-242, Mar 2003.

[27] A. Schäfer and H. Zimmermann, "Recurrent Neural Networks Are Universal Approximators

Artificial Neural Networks – ICANN 2006." vol. 4131, S. Kollias*, et al.*, Eds., ed: Springer Berlin / Heidelberg, 2006, pp. 632-640.

[28] B. Amirikian and H. Nishimura, "What size network is good for generalization of a specific task of interest?," *Neural Networks,* vol. 7, pp. 321-329, 1994.

[29] C. L. G. Steve Lawrence , Ah Chung Tsoi, "What size neural network gives optimal generalization? convergence properties of backpropagation (1996)," 1996.

[30] J. Sietsma and R. J. F. Dow, "Creating artificial neural networks that generalize," *Neural Networks,* vol. 4, pp. 67-79, 1991.

[31] H. Gomi and M. Kawato, "Neural network control for a closed-loop System using Feedback-error-learning," *Neural Networks,* vol. 6, pp. 933-946, 1993.

[32] P. Guo and M. R. Lyu, "A pseudoinverse learning algorithm for feedforward neural networks with stacked generalization applications to software reliability growth data," *Neurocomputing,* vol. 56, pp. 101-121, 2004.

[33] H. T. Huynh and Y. Won, "Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks," *Pattern Recognition Letters,* vol. 32, pp. 1930-1935, 2011.

[34] M. Kimura and R. Nakano, "Learning dynamical systems by recurrent neural networks from orbits," *Neural Networks,* vol. 11, pp. 1589-1599, 1998.

[35] P. Kordík, J. Koutník, J. Drchal, O. Kovářík, M. Čepek, and M. Šnorek, "Meta-learning approach to neural network optimization," *Neural Networks,* vol. 23, pp. 568-582, 2010.

[36] Y. Manabe and B. Chakraborty, "A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network," *Neurocomputing,* vol. 70, pp. 1360-1371, 2007.

[37] M. Masahiko, "Memory and Learning of Sequential Patterns by Nonmonotone Neural Networks," *Neural Networks,* vol. 9, pp. 1477-1489, 1996.

[38] S. Osowski, P. Bojarczak, and M. Stodolski, "Fast Second Order Learning Algorithm for Feedforward Multilayer Neural Networks and its Applications," *Neural Networks,* vol. 9, pp. 1583-1596, 1996.

[39] D. Palmer-Brown and C. Jayne, "Snap–drift neural network for self-organisation and sequence learning," *Neural Networks,* vol. 24, pp. 897-905, 2011.

[40] E. Romero and R. Alquézar, "Comparing error minimized extreme learning machines and support vector sequential feed-forward neural networks," *Neural Networks*.

[41] H. Ugur, "Reinforcement learning in random neural networks for cascaded decisions," *Biosystems,* vol. 40, pp. 83-91, 1997.

[42] H. Ugur, "Reinforcement learning with internal expectation for the random neural network," *European Journal of Operational Research,* vol. 126, pp. 288-307, 2000.

[43] Z.-o. Wang and T. Zhu, "An efficient learning algorithm for improving generalization performance of radial basis function neural networks," *Neural Networks,* vol. 13, pp. 545-553, 2000.

[44] H.-Y. Yu and S.-Y. Bang, "An improved time series prediction by applying the layer-by-layer learning method to FIR neural networks," *Neural Networks,* vol. 10, pp. 1717-1729, 1997.

[45] J. Cao and D. Zhou, "Stability analysis of delayed cellular neural networks," *Neural Networks,* vol. 11, pp. 1601-1605, 1998.

[46] C. Jinde, "On stability of delayed cellular neural networks," *Physics Letters A,* vol. 261, pp. 303-308, 1999.

[47] C. Jinde and W. Jun, "Global asymptotic and robust stability of recurrent neural networks with time delays," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 52, pp. 417-426, 2005.

[48] J. D. Cao and D. W. C. Ho, "A general framework for global asymptotic stability analysis of delayed neural networks based on LMI approach," *Chaos Solitons & Fractals,* vol. 24, pp. 1317-1329, Jun 2005.

[49] C. Feng and R. Plamondon, "On the stability analysis of delayed neural networks systems," *Neural Networks,* vol. 14, pp. 1181-1188, 2001.

[50] M. Forti, S. Manetti, and M. Marini, "Necessary and sufficient condition for absolute stability of neural networks," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on,* vol. 41, pp. 491-494, 1994.

[51] M. Forti, "ON GLOBAL ASYMPTOTIC STABILITY OF A CLASS OF NONLINEAR-SYSTEMS ARISING IN NEURAL-NETWORK THEORY," *Journal of Differential Equations,* vol. 113, pp. 246-264, Oct 1994.

[52] L. Xue-Bin and W. Jun, "An additive diagonal-stability condition for absolute exponential stability of a general class of neural networks," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on,* vol. 48, pp. 1308-1317, 2001.

[53] Z. Jiye, "Globally exponential stability of neural networks with variable delays," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on,* vol. 50, pp. 288-290, 2003.

[54] W. Zheng and J. Zhang, "Global exponential stability of a class of neural networks with variable delays," *Computers & Mathematics with Applications,* vol. 49, pp. 895-902.

[55] C. Wu-Hua and Z. Wei Xing, "Global asymptotic stability of a class of neural networks with distributed delays," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 53, pp. 644-652, 2006.

[56] Z. Zhigang and W. Jun, "Improved conditions for global exponential stability of recurrent neural networks with time-varying delays," *Neural Networks, IEEE Transactions on,* vol. 17, pp. 623-635, 2006.

[57] W. Lisheng and X. Zongben, "Sufficient and necessary conditions for global exponential stability of discrete-time recurrent neural networks," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 53, pp. 1373-1380, 2006.

[58] J. D. E. M. Forti, "On global asymptotic stability of a class of nonlinear systems arising in neural networks theory," *J. Differential Equations,* vol. vol. 113, 1994.

[59] X. Jun, P. Daoying, C. Yong-Yan, and Z. Shouming, "On Stability of Neural Networks by a Lyapunov Functional-Based Approach," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 54, pp. 912-924, 2007.

[60] Z. S. H. Chan, H. W. Ngan, A. B. Rad, A. K. David, and N. Kasabov, "Short-term ANN load forecasting from limited data using generalization learning strategies," *Neurocomputing,* vol. 70, pp. 409-419, 2006.

[61] M. Gori and A. Tesi, "On the problem of local minima in backpropagation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 14, pp. 76-86, 1992.

[62] A. Torn and A. Zilinskas, *Global optimization*: Springer–Verlang, 1987.

[63] S. Yi and B. W. Wah, "Global optimization for neural network training," *Computer,* vol. 29, pp. 45-54, 1996.

[64] K. E. Parsopoulos and M. N. Vrahatis, "On the computation of all global minimizers through particle swarm optimization," *Evolutionary Computation, IEEE Transactions on,* vol. 8, pp. 211-224, 2004.

[65] A. Corana, M. Marchesi, C. Martini, and S. Ridella, "MINIMIZING MULTIMODAL FUNCTIONS OF CONTINUOUS-VARIABLES WITH THE SIMULATED ANNEALING ALGORITHM," *Acm Transactions on Mathematical Software,* vol. 13, pp. 262-280, Sep 1987.

[66] I. Jordanov, "Neural network training and stochastic global optimization," in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, 2002, pp. 488-492 vol.1.

[67] I. Jordanov and A. Georgieva, "Neural Network Learning With Global Heuristic Search," *Neural Networks, IEEE Transactions on,* vol. 18, pp. 937-942, 2007.

[68] M. Zhihong, W. Hong Ren, L. Sophie, and Y. Xinghuo, "A New Adaptive Backpropagation Algorithm Based on Lyapunov Stability Theory for Neural Networks," *Neural Networks, IEEE Transactions on,* vol. 17, pp. 1580-1591, 2006.

[69] S. Ying, C. Zengqiang, and Y. Zhuzhi, "New Chaotic PSO-Based Neural Network Predictive Control for Nonlinear Process," *Neural Networks, IEEE Transactions on,* vol. 18, pp. 595-601, 2007.

[70] W. Cong and D. J. Hill, "Deterministic Learning and Rapid Dynamical Pattern Recognition," *Neural Networks, IEEE Transactions on,* vol. 18, pp. 617-630, 2007.

[71] T. B. Ludermir, A. Yamazaki, and C. Zanchettin, "An Optimization Methodology for Neural Network Weights and Architectures," *Neural Networks, IEEE Transactions on,* vol. 17, pp. 1452-1459, 2006.

[72] E. J. Teoh, K. C. Tan, and C. Xiang, "Estimating the Number of Hidden Neurons in a Feedforward Network Using the Singular Value Decomposition," *Neural Networks, IEEE Transactions on,* vol. 17, pp. 1623-1629, 2006.

[73] Y. Mao, F. Xu-Qian, and L. Xue, "A Class of Self-Stabilizing MCA Learning Algorithms," *Neural Networks, IEEE Transactions on,* vol. 17, pp. 1634-1638, 2006.

[74] H. William W, "Nonlinear principal component analysis of noisy data," *Neural Networks,* vol. 20, pp. 434-443, 2007.

[75] O. Cheol, O. Jun-Seok, and S. G. Ritchie, "Real-time hazardous traffic condition warning system: framework and evaluation," *Intelligent Transportation Systems, IEEE Transactions on,* vol. 6, pp. 265-272, 2005.

[76] D. Srinivasan, J. Xin, and R. L. Cheu, "Evaluation of adaptive neural network models for freeway incident detection," *Intelligent Transportation Systems, IEEE Transactions on,* vol. 5, pp. 1-11, 2004.

[77] K. Thomas and H. Dia, "Comparative evaluation of freeway incident detection models using field data," *Intelligent Transport Systems, IEE Proceedings,* vol. 153, pp. 230-241, 2006.

[78] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic monitoring and accident detection at intersections," in *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*, 1999, pp. 703-708.

[79] H. Dia and G. Rose, "The impact of data quantity on the performance of neural network freeway incident detection models," Monash University. Institute of Transport Studies1996.

[80] F. P. Sioshansi, *Competitive electricity markets: design, implementation, performance*: Elsevier, 2008.

[81] E. Williams and R. A. Rosen, "A BETTER APPROACH

TO MARKET POWER ANALYSIS," *URL: http://www.pulp.tc/tellusmktpwra7-99.pdf.*, 1999.

[82] W. W. Hogan, "COORDINATION FOR COMPETITION IN AN ELECTRICITY MARKET," *Response to an Inquiry Concerning Alternative Power Pooling Institutions Under the Federal Power Act Docket No. RM94-20-000,* March 2, 1995.

[83] H.-P. Chao, S. Oren, and R. Wilson, "Chapter 1 - Reevaluation of Vertical Integration and Unbundling in Restructured Electricity Markets," in *Competitive Electricity Markets*, P. S. Fereidoon, Ed., ed Oxford: Elsevier, 2008, pp. 27-64.

[84] W. W. Hogan, E. G. Read, and B. J. Ring, "Using Mathematical Programming for Electricity Spot Pricing," *International Transactions in Operational Research,* vol. 3, pp. 209-221, 1996.

[85] B. P. Kellerhals, "Pricing Electricity Forwards Under Stochastic Volatility," *SSRN eLibrary,* 2001.

[86] H. Bessembinder and M. L. Lemmon, "Equilibrium Pricing and Optimal Hedging in Electricity Forward Markets," *The Journal of Finance,* vol. 57, pp. 1347-1382, 2002.

[87] J. J. Lucia and E. S. Schwartz, "Electricity Prices and Power Derivatives: Evidence from the Nordic Power Exchange," *Review of Derivatives Research,* vol. 5, pp. 5-50, 2002.

[88] F. A. Longstaff and A. W. Wang, "Electricity Forward Prices: A High-Frequency Empirical Analysis," *The Journal of Finance,* vol. 59, pp. 1877-1900, 2004.

[89] R. Ethier, R. Zimmerman, T. Mount, W. Schulze, and R. Thomas, "A uniform price auction with locational price adjustments for competitive electricity markets," *International Journal of Electrical Power & Energy Systems,* vol. 21, pp. 103-110, 1999.

[90] Mohammad Shahidehpour, Hatim Yamin, and Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*: Wiley-IEEE Press, April 2002.

[91] http://www.pjm.com/ [Online].

[92] F. A. Wolak, "Market design and price behavior in restructured electricity markets: An international comparison," in *Deregulation and Interdependence in the Asia-Pacific Region, NBER-EASE*. vol. 8, T. I. A. O. Krueger, Ed., ed: NBER Books, 2000.

[93] S. L. Puller, "- Pricing and Firm Conduct in California's Deregulated Electricity Market," - *Review of Economics and Statistics. Vol. 89 (1). p 75-87. February 2007.*

[94] S. Vucetic, K. Tomsovic, and Z. Obradovic, "Discovering price-load relationships in California's electricity market," *Power Systems, IEEE Transactions on,* vol. 16, pp. 280-286, 2001.

[95] U. Helman, B. F. Hobbs, and R. P. O'Neill, "Chapter 5 - The Design of US Wholesale Energy and Ancillary Service Auction Markets: Theory and Practice," in *Competitive Electricity Markets*, P. S. Fereidoon, Ed., ed Oxford: Elsevier, 2008, pp. 179-243.

[96] *http://www.recurrence-plot.tk/*.

[97] F. Takens, "Detecting strange attractors in turbulence Dynamical Systems and Turbulence, Warwick 1980." vol. 898, D. Rand and L.-S. Young, Eds., ed: Springer Berlin / Heidelberg, 1981, pp. 366-381.

[98] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena,* vol. 9, pp. 189-208, 1983.

[99] J. J. Healey, "Time series analysis of physical systems possessing homoclinicity," *Phys. D,* vol. 80, pp. 48-60, 1995.

[100] K. Holger, "A robust method to estimate the maximal Lyapunov exponent of a time series," *Physics Letters A,* vol. 185, pp. 77-87, 1994.

[101] N. J. McCullen and P. Moresco, "Method for measuring unstable dimension variability from time series," *Physical Review E,* vol. 73, p. 046203, 2006.

[102] K. Stefanski, K. Buszko, and K. Piecyk, "Transient chaos measurements using finite-time Lyapunov exponents," *Chaos,* vol. 20, Sep 2010.

[103] C. Herui and S. Xiuli, "Research on Electricity Price Forecasting Based on Chaos Theory," in *Future Information Technology and Management Engineering, 2008. FITME '08. International Seminar on*, 2008, pp. 398-401.

[104] L. Zhengjun, Y. Hongming, and L. Mingyong, "Electricity price forecasting model based on chaos theory," in *Power Engineering Conference, 2005. IPEC 2005. The 7th International*, 2005, pp. 1-449.

[105] Y. Hongming and D. Xianzhong, "Chaotic characteristics of electricity price and its forecasting model," in *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, 2003, pp. 659-662 vol.1.

[106] W. Wei, Z. Jian-Zhong, Y. Jing, Z. Cheng-Jun, and Y. Jun-Jie, "Prediction of spot market prices of electricity using chaotic time series," in *Machine Learning and*

Cybernetics, 2004. Proceedings of 2004 International Conference on, 2004, pp. 888-893 vol.2.

[107] S. El Boustani and A. Destexhe, "BRAIN DYNAMICS AT MULTIPLE SCALES: CAN ONE RECONCILE THE APPARENT LOW-DIMENSIONAL CHAOS OF MACROSCOPIC VARIABLES WITH THE SEEMINGLY STOCHASTIC BEHAVIOR OF SINGLE NEURONS?," *International Journal of Bifurcation and Chaos,* vol. 20, pp. 1687-1702, Jun 2010.

[108] J. B. Gao, S. K. Hwang, and J. M. Liu, "When can noise induce chaos?," *Physical Review Letters,* vol. 82, pp. 1132-1135, Feb 1999.

[109] S. K. Hwang, J. B. Gao, and J. M. Liu, "Noise-induced chaos in an optically injected semiconductor laser model," *Physical Review E,* vol. 61, pp. 5162-5170, May 2000.

[110] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE,* vol. 78, pp. 1550-1560, 1990.

[111] B. A. Pearlmutter, "Gradient calculations for dynamic recurrent neural networks: a survey," *Neural Networks, IEEE Transactions on,* vol. 6, pp. 1212-1228, 1995.

[112] G. Serpen and Y. Xu, "Weight Initialization for Simultaneous Recurrent Neural Network Trained with A Fixed-Point Learning Algorithm," *Neural Processing Letters,* vol. 17, pp. 33-41, 2003.

[113] D.-Z. Du and X.-S. Zhang, "Global convergence of Rosen's gradient projection method," *Mathematical Programming,* vol. 44, pp. 357-366, 1989.

[114] J. a. O. Lee, J, "Hybrid learning of mapping and its Jacobian in multilayer neural networks," *Neural Computation,* vol. Vol. 9, pp. 937-958, July 1, 1997.

[115] D. Chelidze and M. Liu, "Reconstructing slow-time dynamics from fast-time measurements," *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences,* vol. 366, pp. 729-745, Mar 2008.

[116] Chelidze D and C. J. P, "Phase space warping: nonlinear time-series analysis for slowly drifting systems," *Phil. Trans. R. Soc. A,* vol. 364, pp. 2495-2513, 2006.

[117] C. Lucheroni, "Resonating models for the electric power market," *Physical Review E,* vol. 76, p. 056116, 2007.

[118] F. C. Hoppensteadt and E. M. Izhikevich, *Weakly Connected Neural Networks*.Applied Mathematical Sciences vol. 126: Springer, 1997.

[119] H.-G. Beyer, *The Theory of Evolution Strattegies*.Natural Computing vol. XIX: Springer, 2001.

[120] H. Zareipour, K. Bhattacharya, and C. A. Canizares, "Forecasting the hourly Ontario energy price by multivariate adaptive regression splines," in *2006 IEEE Power Engineering Society General Meeting, 18-22 June 2006*, Piscataway, NJ, USA, 2006, p. 7 pp.

[121] H. Zareipour, C. A. Canizares, K. Bhattacharya, and J. Thomson, "Application of public-domain market information to forecast Ontario's wholesale electricity prices," *IEEE Transactions on Power Systems,* vol. 21, pp. 1707-17, 2006.

[122] D. Mirikitani and N. Nikolaev, "Nonlinear maximum likelihood estimation of electricity spot prices using recurrent neural networks," *Neural Computing & Applications,* 2010.

[123] A. M. Gonzalez, A. M. S. Roque, and J. Garcia-Gonzalez, "Modeling and forecasting electricity prices with input/output hidden Markov models," *IEEE Transactions on Power Systems,* vol. 20, pp. 13-24, 2005.

[124] C. P. Rodriguez and G. J. Anders, "Energy price forecasting in the Ontario competitive power system market," *IEEE Transactions on Power Systems,* vol. 19, pp. 366-74, 2004.

[125] J. P. S. Catalao, H. M. I. Pousinho, and V. M. F. Mendes, "Neural networks and wavelet transform for short-term electricity prices forecasting," in *2009 15th International Conference on Intelligent System Applications to Power Systems (ISAP), 8-12 Nov. 2009*, Piscataway, NJ, USA, 2009, p. 5 pp.

[126] S. Max, "Filtering and Forecasting Spot Electricity Prices in the Increasingly Deregulated Australian Electricity Market," Quantitative Finance Research Centre, University of Technology, SydneySep 2001.

[127] N. M. Pindoriya, S. N. Singh, and S. K. Singh, "An Adaptive Wavelet Neural Network-Based Energy Price Forecasting in Electricity Markets," *Power Systems, IEEE Transactions on,* vol. 23, pp. 1423-1432, 2008.

[128] P. Areekul, T. Senjyu, H. Toyama, and A. Yona, "A hybrid ARIMA and neural network model for short-term price forecasting in deregulated market," *IEEE Transactions on Power Systems,* vol. 25, pp. 524-30, 2010.

[129] D. Srinivasan, Y. Fen Chao, and L. Ah Choy, "Electricity Price Forecasting Using Evolved Neural Networks," in *Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on*, 2007, pp. 1-7.

[130] Z. Jun Hua, D. Zhao Yang, X. Zhao, and W. Kit Po, "A statistical approach for interval forecasting of the electricity price," *IEEE Transactions on Power Systems,* vol. 23, pp. 267-76, 2008.

[131] G. Papadopoulos, P. J. Edwards, and A. F. Murray, "Confidence estimation methods for neural networks: a practical comparison," *IEEE Transactions on Neural Networks,* vol. 12, pp. 1278-87, 2001.

[132] Nandram Balgobin, J. D. Petruccelli, and C. Minghui, *Applied statistics for engineers and scientists*, 1st ed.: Prentice Hall, 1999.

[133] P. F. Christoffersen, "Evaluating Interval Forecasts," *International Economic Review* vol. 39, pp. 841-862, 1997.

[134] K. F. Wallis, "Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts," *International Journal of Forecasting,* vol. 19, pp. 165-75, 2003.

[135] Y. Liu, "Value-at-Risk Model Combination Using Artificial Neural Networks," Emory University Working Paper Series2005.

[136] J. Guckenheimer and M. D. LaMar, "Periodic Orbit Continuation in Multiple Time Scale Systems Numerical Continuation Methods for Dynamical Systems," B. Krauskopf*, et al.*, Eds., ed: Springer Netherlands, 2007, pp. 253-267.

[137] K. Al-Naimee, S. F. Abdalah, M. Ciszak, R. Meucci, F. T. Arecchi, and F. Marino, "Incomplete Homoclinic Scenarios in Semiconductor Devices with Optoelectronic Feedback: Generation and Synchronization," in *Complexity in Engineering, 2010. COMPENG '10.*, 2010, pp. 109-111.

[138] E. M. Izhikevich, "Neural Excitability, Spiking, and Bursting," *International Journal of Bifurcation and Chaos* vol. 10, pp. 1171--1266, 2000.

[139] V. O. Khavrus, P. E. Strizhak, and A. L. Kawczynski, "Scalings of mixed-mode regimes in a simple polynomial three-variable model of nonlinear dynamical systems," *Chaos,* vol. 13, pp. 112-122, Mar 2003.

[140] Shilnikov L.P., Shilnikov A., Turaev D., and L. Chua, *Methods of Qualitative Theory in Nonlinear Dynamics. Part I* World Sci., 1998.

[141] F. Marino, F. Marin, S. Balle, and O. Piro, "Chaotically Spiking Canards in an Excitable System with 2D Inertial Fast Manifolds," *Physical Review Letters,* vol. 98, p. 074104, 2007.

[142] M. T. M. Koper, P. Gaspard, and J. H. Sluyters, "A one-parameter bifurcation analysis of the indium/thiocyanate electrochemical oscillator," *The Journal of Physical Chemistry,* vol. 96, pp. 5674-5675, 1992/07/01 1992.

[143] K. Christian, "A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics," *Physica D: Nonlinear Phenomena,* vol. 240, pp. 1020-1035, 2011.

[144] L. Arnold, C. Jones, K. Mischaikow, and G. Raugel, "Geometric singular perturbation theory Dynamical Systems." vol. 1609, ed: Springer Berlin / Heidelberg, 1995, pp. 44-118.

[145] J. Grasman, *Asymptotic Methods for Relaxation Oscillations and Applications* vol. Vol. 63: Springer, 1987.

[146] K. J. Hunt, D. Sbarbaro, R. Żbikowski, and P. J. Gawthrop, "Neural networks for control systems—A survey," *Automatica,* vol. 28, pp. 1083-1112, 1992.

[147] A. MeyerBase, F. Ohl, and H. Scheich, "Singular perturbation analysis of competitive neural networks with different time scales," *Neural Computation,* vol. 8, pp. 1731-1742, Nov 1996.

[148] A. Meyer-Base, F. Ohl, and H. Scheich, "Quadratic-type Lyapunov functions for competitive neural networks with different time-scales," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 1995, pp. 3210-3214 vol.6.

[149] A. Meyer-Baese, "Flow invariance for competitive neural networks with different time-scales," in *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on*, 2002, pp. 858-861.

[150] A. Meyer-Baese, S. S. Pilyugin, and Y. Chen, "Global exponential stability of competitive neural networks with different time scales," *Neural Networks, IEEE Transactions on,* vol. 14, pp. 716-719, 2003.

[151] A. Meyer-Base, S. S. Pilyugin, and A. Wismuller, "Stability analysis of a self-organizing neural network with feedforward and feedback dynamics," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 2004, pp. 1505-1509 vol.2.

[152] A. Meyer-Base, F. Ohl, and H. Scheich, "Stability analysis techniques for competitive neural networks with different time-scales," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 1995, pp. 3215-3219 vol.6.

[153] W. Y. Liu, "Neural adaptive control for nonlinear multiple time scale dynamic systems," *Available: http://ctrl.cinvestav.mx/~yuw/*.

[154] A. C. Sandoval, Y. Wen, and L. Xiaoou, "Some stability properties of dynamic neural networks with different time-scales," in *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, 2006, pp. 4218-4224.

[155] D. H. Rao, M. M. Gupta, and H. C. Wood, "Neural networks in control systems," in *WESCANEX 93. 'Communications, Computers and Power in the Modern Environment.' Conference Proceedings., IEEE*, 1993, pp. 282-290.

[156] S. Grossberg, "ADAPTIVE PATTERN-CLASSIFICATION AND UNIVERSAL RECODING .1. PARALLEL DEVELOPMENT AND CODING OF NEURAL FEATURE DETECTORS," *Biological Cybernetics,* vol. 23, pp. 121-134, 1976.

[157] S. Grossberg, "ADAPTIVE PATTERN-CLASSIFICATION AND UNIVERSAL RECODING .2. FEEDBACK, EXPECTATION, OLFACTION, ILLUSIONS," *Biological Cybernetics,* vol. 23, pp. 187-202, 1976.

[158] R. O'Malley, "A singular perturbations approach to reduced-order modeling and decoupling for large scale linear systems Numerical Integration of Differential Equations and Large Linear Systems." vol. 968, J. Hinze, Ed., ed: Springer Berlin / Heidelberg, 1982, pp. 246-255.

[159] H. Zareipour, K. Bhattacharya, and C. A. Canizares, "Forecasting the hourly Ontario energy price by multivariate adaptive regression splines," in *Power Engineering Society General Meeting, 2006. IEEE*, 2006, p. 7 pp.

[160] A. De Sanctis and C. Mari, "Modelling spikes in electricity markets using excitable dynamics," *Physica A: Statistical Mechanics and its Applications,* vol. 384, pp. 457-467, 2007.

[161] S. K. Aggarwal, L. M. Saini, and A. Kumar, "Electricity price forecasting in Ontario electricity market using wavelet transform in artificial neural network based model," *International Journal of Control Automation and Systems,* vol. 6, pp. 639-650, Oct 2008.

[162] S. K. Aggarwal, L. M. Saini, and A. Kumar, "Price forecasting using wavelet transform and LSE based mixed model in Australian electricity market," *International Journal of Energy Sector Management,* vol. Vol. 2, pp. 521 - 546, 2008.