J. Appl. Environ. Biol. Sci., 4(7S)99-106, 2014

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

A Review of Adaptive Bayesian Modeling for Time Series Forecasting

Tahseen Jilani and Muhammad Najamuddin

Assistant Professor, Department of Computer Science, University of Karachi,
Lecturer, Department of Statistics, Federal Urdu University of Arts, Sciences & Technology, Karachi
Received: September 1, 2014
Accepted: November 13, 2014

ABSTRACT

Statistical time series forecasting is one of the fruitful areas having applications ranging from finance, economics, risk management, environmental sciences, geo sciences, genomics and etc. In time series forecasting models, Bayesian Modeling is one of the most successful areas having applications in time series parameter estimation, apriori relationship modeling and causality analysis in d-separated or free conditions. Due to heavy iterative differential equations, boundary conditions, and incomplete integral problems, scientists have started using derivative free and approximation algorithms that are adaptive in nature and have shown better performances over conventional statistical time series models, these adaptive models include artificial neural networks, genetic algorithms, evolutionary algorithms, nature inspired algorithms and etc. These adaptive models has proven results in different domains of applications including time series forecasting, but the field of adaptive models using Bayesian approaches for time series forecasting is an immature fields. In this paper, a review on the adaptive Bayesian modeling is presented. We show how Bayesian framework is used in neural network, genetic algorithms, wavelet analysis for time series analysis and forecasting. We discuss the applications of Bayesian approach with different methods for various time series data including wind, hydrological, financial, rainfall, electricity load time series data etc. It is shown that there is big room for new research work and have proposed some new approaches as well.
KEYWORDS: Bayesian framework, neural networks, genetic algorithms, wavelets, statistical time series modeling and forecasting.

1 INTRODUCTION

For decades, scientists and engineers are struggling to find the underlying structure of a random phenomenon normally called kernel. In statistical terms, this is called kernel estimation or probability model estimation for some underlying problem. There are a number of techniques in parametric, non-parametric and semi- parametric approaches that are used for this purpose [1]. In Bayesian modeling, we can foresee any random process based on the prior information. Thus for Bayesian data analyses approaches, we require a well-defined prior data structure so that be able to construct the posterior probability distribution/ structure for the unobserved variable(s) based on the prior information. Thus, in every field where a causal relationship holds true, we can adapt Bayesian approaches to model the situation.

The Bayesian neural networks (BNNs) are used for modeling the nonlinear time series because of their competencies to manage with the intricacy issue. Moreover, a natural interpretation of the estimation and the forecasts from the estimated models are ensured by them. Due to this purpose, BNNs are used with the time series, the regression, the classification and the density estimation problems. Basically, Gaussian approximation, the ensemble learning and Markov Chain Monte Carlo (MCMC) simulations are the typical base of Bayesian treatments of the learning in the ANNs, known as the full Bayesian approach.

For ANNs, [2] and [3] introduced Gaussian approximation known as Laplace's method. The approach of modeling the posterior distribution by a Gaussian distribution, centered locally at a mode of posterior distribution of parameters [4]. [5] introduced the ensemble learning, in which the fitting of approximating distribution can be done by minimizing a Kullback–Leibler divergence globally rather than locally. In full Bayes framework, advanced Bayesian simulation methods is introduced by [6], in which

*** Corresponding Author:** Tahseen Jilani, Assistant Professor, Department of Computer Science, University of Karachi, tahseenjilani@uok.edu.pk

the parameter samples from the posterior distribution can be generated by using MCMC simulations. Bayesian learning is integrated with Hybrid Monte Carlo (HMC) method by [6] due to computational expensiveness of MCMC techniques, and also suffering from measuring the convergence. Later, [7], [4] and [8] reviewed the Bayesian applications to ANNs in detail.

2 Adaptive Models.

2.1 Neural Networks

Neural networks are parameterized nonlinear models used for empirical regression. Neural networks are built out of interconnected neurons (or units) arranged in layers. The network's architecture used for this work is the perceptron made of three layers [7]; the input layer, one hidden layer and the output layer.

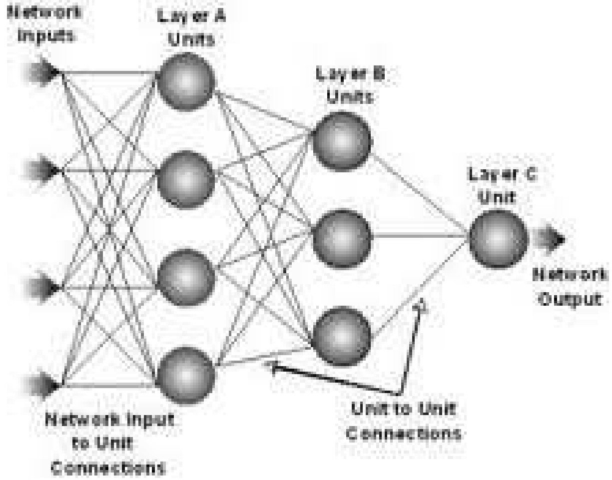

Figure 1: A simple artificial neural network model

In the literature, some remarkable studies are there, in which focus is specifically made on the ANNs from Bayesian perspective related problem. For example; for the dimensionality problem in the ANNs, the issue of selecting the number of hidden neurons with the growing and the pruning algorithms have been discussed by [9], [10] and [11]. The sequential Monte Carlo (MC) methods and the particle filters are incorporated in the BNNs by [12]. [13] proposed an MCMC algorithm with Hierarchical Bayesian approach for nonlinear time series prediction. They used predictive distributions for drawing samples for the computation of predictive mean and error bar. The algorithm is tested by (chaotic) R¨ossler system generated time series and quadratic approximations are outperformed by it.

[14] proposed a novel neural network method with Hierarchical Bayesian method for nonlinear time series forecasting. The approach is applied on power demand forecasting for buildings and chaotic time series data. [15] introduced a Bayesian feed forward neural network prediction model for time series data. An inference scheme based on MCMC simulation is presented. To deal with model selection, we analyze why standard birth/death reversible jump moves result in slowly mixing MCMC. Instead, we propose new reversible jump moves to add or delete special kinds of nodes characterized as linearized, irrelevant or duplicate. For illustration of the methodology they used the airline passengers and the lynx time series data. [16] applied BNN models to predict motor vehicle crashes. They compared back propagation neural network (BPNN), Bayesian neural network (BNN) and the negative binomial (NB) regression models. The techniques are applied on collected data from Texas. They showed that in forecasting comparison NB regression model outperformed by both neural network models in general. Also BNN model gives better abilities to generalize than the BPNN model. And the over-fitting problem can be alleviated effectively

even nonlinear approximation ability would not be compromised significantly. Also showed in the safety of highway other valuable analysis could be done by BNN.

[17] forecast short term electric load by design a neural network model based on Bayesian approach. They applied the Bayesian neural network to real load data. Forecasting accuracy is assessed by MAPE and RMSE. They showed that the proposed model can efficiently deal the model complexity as compared to cross validation NN techniques by using the evidence framework and model selection.

[18] forecast electricity prices by proposing a BNN model. Some remarkable advantages are presented over the classical NN models. The network over-fitting can be avoided, the uncertainty degree in the forecasts can be indicated, an appropriate scale can be automatically selected for network weights and, consequently, the optimal model for prediction can be selected with the proposed BNN model. Also they found superior results in the comparison of proposed method with ARIMA, Wavelet-ARIMA and a classical neural network method.

[19] proposed a time-delay back propagation neural network model on the basis of Bayesian regularization. They forecasted the import and export trades in one industry with the model. They resulted that the proposed model capable in generalization efficiently. Also the historical curve can be learned and business trends can be predicted efficiently. [20] evaluated automatic NN modeling techniques using Bayesian approach. They applied the model to daily load and weather data for six samples of four different countries. The analysis of Bayesian ARD input selection is made, and the best structure selection by Bayesian approach as compared to cross validation method is described. They also compared the Bayesian results to cross validation and naïve based NN method.

[21] proposed an adaptive forecast method for very short term prediction of wind power. An ANN model is employed together with adaptive Bayesian learning and Gaussian process approximation. It is established for two cases the proposed method outperforms the persistent method. [22] presented a robust two-step method to forecast wind speed. The proposed method is obtained by applying Bayesian combination algorithm and ADALINE, BP and RBF neural network models for forecasting. They resulted that reliable, adaptive and relatively precise predictions can be obtained by the Bayesian combination method.

[23] presented a model based on BNN learned by the Hybrid Monte Carlo algorithm to forecast electricity load. Laplace and HMC algorithm is used for BNN and back propagation algorithm is used for ANN. They forecast 25 days hourly load of four seasons each. Forecast evaluation was measured by MAPE and RMSE. They showed that hybrid Monte Carlo BNNs model outperform the other model and the over-fitting problem can be solved from it due to powerful capability of generalization.

[24] proposed Bayesian neural network (BNN) with adjusting procedures. They compared the proposed model with natural BNN and feed forward neural network (FFNN). Mean square error (MSE) is used for forecasting accuracy. They used rainfall data in Jatisrono station, central Java, Indonesia. They showed that BNN with adjusting procedure outperforms the natural BNN and FFNN. [25] proposed a novel artificial neural network with Bayesian regularization for financial market forecasting. Goldman Sachs Group Inc. and Microsoft Corp. stock are used for determination of the model effectiveness. They resulted that the model outperforms the ARIMA and the fusion model without need for data preprocessing, test of seasonality, or analysis of cycle.

2.2 Genetic Algorithm

GAs developed by John Holland in 1970's are heuristic optimization methods based on the concepts of natural evolution, and belongs to the larger class of evolutionary algorithms [26] and [27]. They consist of the artificial operators such as selection, mutation, crossover and migration that are the components of the natural evaluation process. In the selection process of GAs, the possibilistic (or linguistic) and the probabilistic uncertainties arise when the parents (the parameter vectors), who would create the next generation, are being selected into the mating pool [28]and [29].

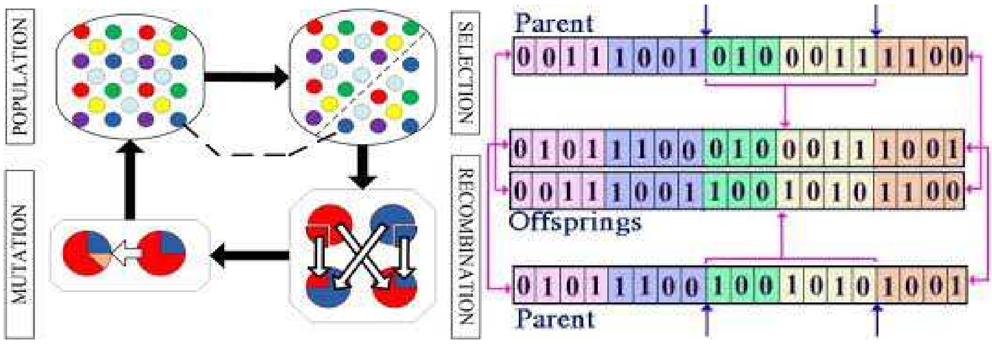

Figure 2: (a). The four operations of a simple Gnetic Algorithm **(b).** An example of GA application to binary population

[30] proposed to the evolutionary MC algorithm which samples the parameters in the ANNs from the Boltzmann distribution using the mutation, the crossover and the exchange operations defined in the genetic algorithms (GAs). [31] adapted to the mutation and the crossover operators defined in GAs into Bayesian learning, and estimated the parameters using Genetic MC algorithm. [28] integrated the hierarchical Bayesian learning with GAs and the fuzzy numbers to estimate the parameters in the ANNs.[32] proposed an evolutionary Monte Carlo (MC) algorithm is proposed to train the Bayesian neural networks (BNNs) for the time series forecasting. This approach called as Genetic MC is based on Gaussian approximation with recursive hyper-parameter. Genetic MC integrates MC simulations with the genetic algorithms and the fuzzy membership functions. In the implementations, Genetic MC is compared with the traditional neural networks and time series techniques in terms of their forecasting performances over the weekly sales of a Finance Magazine. [29] proposed a two phased method, where the length of fuzzy intervals were estimated using genetic algorithms and the time series forecasting is performed using a simple PSO algorithm.

3 TIME SERIES METHODS

In the literature, there are many stochastic processes described over time. These processes have quantities related to recent time periods influenced by their past values. This structure is based on time series methodology. The approach of [33] is given as a classic reference on time series techniques in order to model the functional structure [34]. [35] presented a nonlinear Bayesian multivariate adaptive regression model to forecast average wind speeds which was collected at several different sites in a complex terrain, where only a short run of data is available.

[36]evaluated the forecasting precision of linear autoregressive, smooth transition autoregressive (STAR), and neural network (NN) time series models for 47 monthly macro-economic variables of the G7 economies. An NN model is obtained using Bayesian regularization. They resulted that generally linear autoregressive models are outperformed by STAR model. The NN models provide mix results specifically at long horizons forecasting, more accurate forecasts are produced by Bayesian NN model than a corresponding model specified using the specific-to-general approach.

[37] proposed a Monte Carlo algorithm for Bayesian neural networks (BNNs) training. Despite hidden units, a prior on network connections is put in BNN model selection. The BNN model is compared with ARIMA, bilinear, TAR, and traditional neural network model, on various data sets. He showed that the BNN model outperforms the other models. And generalization capability of BNNs is discovered in many respects.

[38] explored chaos theory as a tool from non-linear time series analysis to automatic select the lags of the load series data that will be used by the neural models. Bayesian inference applied to MLPs

and relevance vector machines are used in the development of autonomous neural models. They showed that the results are comparable especially for the BIAMLPs, show the potential of the proposal.

[39] proposed a new homogeneous neural network ensemble approach called Generalized Regression Neural Network (GEFTS–GRNN) Ensemble for Forecasting Time Series. GEFTS uses a dynamic nonlinear weighting system wherein the outputs from several base-level GRNNs are combined using a combiner GRNN to produce the final output. 11 most used algorithms are compared with GEFTS on 30 real datasets. They found that the proposed algorithm performs efficiently than existing ones. Also GEFTS effectively forecast seasonally patterned time series. [40] proposed a new fuzzy time series model for multivariate fuzzy time series using fuzzy logical relationship groups [41].

[42] compared ANNs and adaptive neural-based fuzzy inference system (ANFIS) with autoregressive (AR) model to forecast monthly reservoir inflow. Flow time series data of the Sutlej River at Bhakra Dam, India is applied. In forecasting comparison ANFIS outperform the AR and ANN models in all cases. The cyclic terms is incorporated with the proposed ANFIS model, which give improved representation of the monthly inflow forecasting for planning and operation of reservoir.

4 Wavelet

Wavelet decomposition provides a way of analyzing a signal in both time and frequency domains. Wavelets have been found to provide an effective model for data of the form y = f + z, when f is a potentially complex, spatially inhomogeneous function. The essence of a wavelet-based model is a one-to-one transform off into a space of wavelet coefficients. The coefficient space is structured roughly according to the location and scale (frequency) of the functional information contained in each coefficient. Standard wavelet methods assume equally spaced measurements off with additive noise, and seek to "denoise" the data by shrinking the empirical wavelet coefficients toward O.When the reduced empirical coefficients are then transformed back to the data space, the reconstructed signal typically has much of the noise removed [43]and [44].

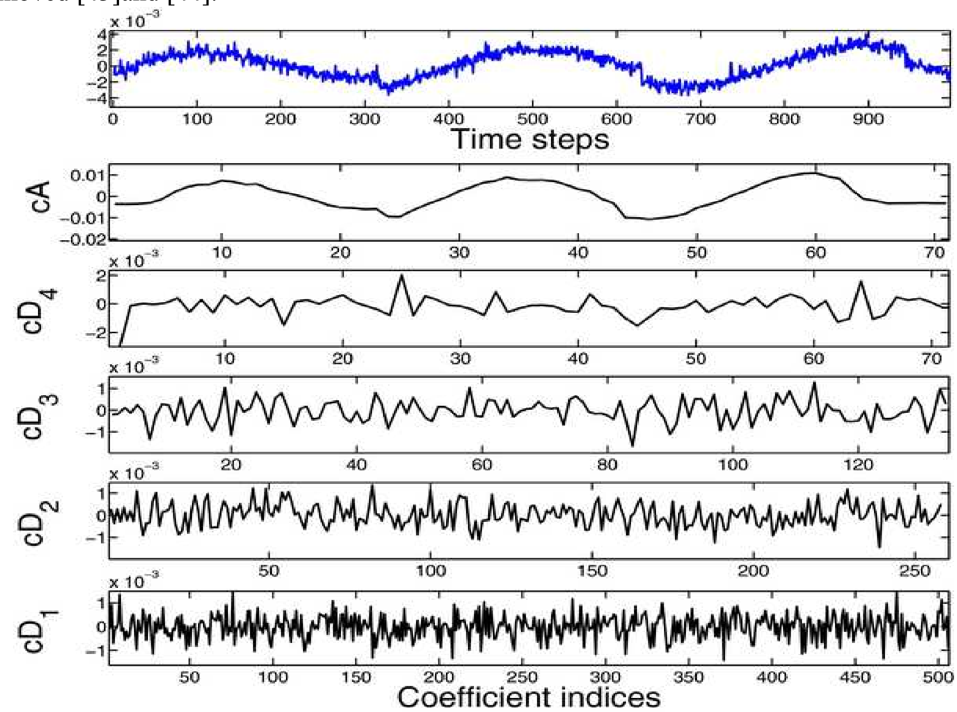

Figure 3. Example of a discrete wavelet transform (DWT). Top: sinusoidal time series with Gaussian noise and saw-tooth functions superimposed

[45] proposed a novel neural wavelet model to forecast electricity load. The forecast model is based on wavelet multi-resolution decomposition by autocorrelation shell representation and neural networks (multilayer perceptrons, or MLPs) modeling of wavelet coefficients. The practical Bayesian method ARD model is applied to choose the size of MLPs, which are then trained to provide forecasts. The proposed techniques have been tested with the Australian electricity market data series with promising results. Different approaches for load forecast are also performed on the same load series for comparison to show the effectiveness of the proposed hybrid forecast model. [46] employed the adaptive Metropolis-Markov chain Monte Carlo (AM-MCMC) algorithm to wavelet regressive modeling processes (called AM-MCMC-WR) was proposed for hydrologic time series forecasting. The parameters' uncertainty is estimated by AM-MCMC algorithm in WR model, which is the base to provide probabilistic hydrologic time series forecast. They resulted that the forecasting performance of AM-MCMC-WR and WR models are identical but outperform the linear regression models.

5. Conclusions and Future Studies

The statistical approaches for parameter estimation work under many assumptions. On the other hand, Machine learning algorithms have proven applications in pattern classification and regression modeling and so the objective of hyper-parameters' estimation are carried out using machine learning algorithms. These methods are superior to statistical hyper-parameters' estimation methods because we don't need to follow statistical assumptions like normality and stationairty. Furthermore, derivative free methods and nature inspired algorithms optimize the objective function without assuming any differential equation although they are iterative too. Thus time and space complexity issues are dramatically reduced using machine learning approaches. To overcome the excessive parameters in Bayesian estimation, the complexity is highly reducible.

As the research in machine learning algorithms is one of the top research areas, therefore, the development of new machine learning algorithms is an open challenge. On the other side, when using these ML approaches for Bayesian estimation, then the hybrid models will result in better estimation with lower complexity.

We aim to develop new adaptive Bayesian estimation models by using low complexity and cheaper algorithms from machine learning areas.

REFERENCES

1 Duda, R.O., Hart, P.E., and Stork, D.G.: 'Pattern classification' (John Wiley & Sons, 2012. 2012)
2 Buntine, W.L., and Weigend, A.S.: 'Computing second derivatives in feed-forward networks: A review', Neural Networks, IEEE Transactions on, 1994, 5, (3), pp. 480-488
3 MacKay, D.J.: 'A practical Bayesian framework for backpropagation networks', Neural computation, 1992, 4, (3), pp. 448-472
4 MacKay, D.J.: 'Probable networks and plausible predictions-a review of practical Bayesian methods for supervised neural networks', Network: Computation in Neural Systems, 1995, 6, (3), pp. 469-505
5 Hinton, G.E., and Van Camp, D.: 'Keeping the neural networks simple by minimizing the description length of the weights', in Editor (Ed.)^(Eds.): 'Book Keeping the neural networks simple by minimizing the description length of the weights' (ACM, 1993, edn.), pp. 5-13
6 Neal, R.M.: 'Bayesian training of backpropagation networks by the hybrid Monte Carlo method', in Editor (Ed.)^(Eds.): 'Book Bayesian training of backpropagation networks by the hybrid Monte Carlo method' (Citeseer, 1992, edn.), pp.
7 Bishop, C.M.: 'Neural networks for pattern recognition', 1995
8 Neal, R.M.: 'Bayesian learning for neural networks', University of Toronto, 1995
9 Insua, R., and Muller, P.: 'Feedforward neural networks for nonparametric regression', in Editor (Ed.)^(Eds.): 'Book Feedforward neural networks for nonparametric regression' (Institute of statistics and decision sciences, Duke University, 1998, edn.), pp.

10 Holmes, C., and Mallick, B.: 'Bayesian radial basis functions of variable dimension', Neural computation, 1998, 10, (5), pp. 1217-1233

11 Marrs, A.D.: 'An application of reversible-jump MCMC to multivariate spherical Gaussian mixtures', Advances in neural information processing systems, 1998, pp. 577-583

12 de Freitas, J.F., Niranjan, M., and Gee, A.H.: 'Dynamic learning with the EM algorithm for neural networks', Journal of VLSI signal processing systems for signal, image and video technology, 2000, 26, (1-2), pp. 119-131

13 Nakada, Y., Kurihara, T., and Matsumoto, T.: 'Bayesian MCMC nonlinear time series prediction', in Editor (Ed.)^(Eds.): 'Book Bayesian MCMC nonlinear time series prediction' (IEEE, 2001, edn.), pp. 3509-3512

14 Matsumoto, T., Hamagishi, H., Sugi, J., Saito, M., and Chonan, Y.: 'From data to nonlinear dynamics: Time series prediction via hierarchical Bayes approach', Electronics and Communications in Japan (Part II: Electronics), 2002, 85, (10), pp. 71-83

15 Menchero, A., Diez, R.M., Insua, D.R., and Müller, P.: 'Bayesian analysis of nonlinear autoregression models based on neural networks', Neural computation, 2005, 17, (2), pp. 453-485

16 Xie, Y., Lord, D., and Zhang, Y.: 'Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis', Accident Analysis & Prevention, 2007, 39, (5), pp. 922-933

17 Lauret, P., Fock, E., Randrianarivony, R.N., and Manicom-Ramsamy, J.-F.: 'Bayesian neural network approach to short time load forecasting', Energy conversion and management, 2008, 49, (5), pp. 1156-1166

18 Vahidinasab, V., and Jadid, S.: 'Bayesian neural network model to predict day-ahead electricity prices', European Transactions on Electrical Power, 2010, 20, (2), pp. 231-246

19 Jiang, W., Zhang, L., and Wang, P.: 'Nonlinear time series forecasting of time-delay neural network embedded with Bayesian regularization', Applied Mathematics and Computation, 2008, 205, (1), pp. 123-132

20 Hippert, H.S., and Taylor, J.W.: 'An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting', Neural networks, 2010, 23, (3), pp. 386-395

21 Blonbou, R.: 'Very short-term wind power forecasting with neural networks and adaptive Bayesian learning', Renewable Energy, 2011, 36, (3), pp. 1118-1124

22 Li, G., Shi, J., and Zhou, J.: 'Bayesian adaptive combination of short-term wind speed forecasts from neural network models', Renewable Energy, 2011, 36, (1), pp. 352-359

23 Niu, D.-x., Shi, H.-f., and Wu, D.D.: 'Short-term load forecasting using bayesian neural networks learned by Hybrid Monte Carlo algorithm', Applied Soft Computing, 2012, 12, (6), pp. 1822-1827

24 Fithriasari, K., Iriawan, N., Ulama, B., Sutikno, and Kuswanto, H.: 'Prediction of Hourly Rainfall using Bayesian Neural Network with Adjusting Procedure', The Third Basic Science International Conference – 2013, 2013

25 Ticknor, J.L.: 'A Bayesian regularized artificial neural network for stock market forecasting', Expert Systems with Applications, 2013, 40, (14), pp. 5501-5506

26 Goldberg, D.E., and Holland, J.H.: 'Genetic algorithms and machine learning', Machine learning, 1988, 3, (2), pp. 95-99

27 Holland, J.H.: 'Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence' (U Michigan Press, 1975. 1975)

28 Kocadag˘li, O.: 'Hybrid Bayesian neural networks with genetic algorithms and fuzzy membership functions', Mimar Sinan F.A. University, 2012

29 Jilani, T., Nikos, M., and Amjad, U.: 'A Hybrid Genetic Algorithm and Particle Swarm Optimization Based Fuzzy Time Series Model TAIFEX and KSE-100 Forecasting'. Proc. NAUN-First International Confernce on Biological Inspired Computation, University of Algaro, Portugal, 22-24, May-2012 2012 pp. Pages

30 Liang, F., and Wong, W.H.: 'Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models', Journal of the American Statistical Association, 2001, 96, (454), pp. 653-666

31 Marwala, T., and Crossingham, B.: 'Neuro-rough models for modelling HIV', in Editor (Ed.)^(Eds.): 'Book Neuro-rough models for modelling HIV' (IEEE, 2008, edn.), pp. 3089-3095

32 Kocadağlı, O., and Aşıkgil, B.: 'Nonlinear time series forecasting with Bayesian neural networks', Expert Systems with Applications, 2014

33 Box, G.E., Jenkins, G.M., and Reinsel, G.C.: 'Time series analysis: forecasting and control' (John Wiley & Sons, 2013. 2013)

34 Davidson, R., and MacKinnon, J.G.: 'Estimation and inference in econometrics', OUP Catalogue, 1993

35 Denison, D., Dellaportas, P., and Mallick, B.: 'Wind speed prediction in a complex terrain', Environmetrics, 2001, 12, (6), pp. 499-515

36 Teräsvirta, T., Van Dijk, D., and Medeiros, M.C.: 'Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination', International Journal of Forecasting, 2005, 21, (4), pp. 755-774

37 Liang, F.: 'Bayesian neural networks for nonlinear time series forecasting', Statistics and Computing, 2005, 15, (1), pp. 13-29

38 Ferreira, V.H., and Silva, A.P.A.d.: 'Chaos theory applied to input space representation of autonomous neural network-based short-term load forecasting models', Sba: Controle & Automação Sociedade Brasileira de Automatica, 2011, 22, (6), pp. 585-597

39 Gheyas, I.A., and Smith, L.S.: 'A novel neural network ensemble architecture for time series forecasting', Neurocomputing, 2011, 74, (18), pp. 3855-3864

40 Jilani, T.A., and Burney, S.M.A.: 'Multivariate stochastic fuzzy forecasting models', Expert Systems with Applications, 2008, 35, (3), pp. 691-700

41 Amjad, U., Jilani, T., and F, Y.: 'A Two Phase Algorithm for Fuzzy Time Series Forecasting Using Genetic Algorithm and Particle Swarm Optimization Techniques'. Proc. International Journal of Computer Applications (IJCA)2012 pp. Pages

42 Lohani, A., Kumar, R., and Singh, R.: 'Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques', Journal of Hydrology, 2012, 442, pp. 23-35

43 DeVore, R., and Lucier, B.: 'Fast wavelet techniques for near-optimal image processing', in Editor (Ed.)^(Eds.): 'Book Fast wavelet techniques for near-optimal image processing' (IEEE, 1992, edn.), pp. 48.43-48.43

44 Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., and Picard, D.: 'Wavelet shrinkage: asymptopia?', Journal of the Royal Statistical Society. Series B (Methodological), 1995, pp. 301-369

45 Zhang, B.-L., and Dong, Z.-Y.: 'An adaptive neural-wavelet model for short term load forecasting', Electric Power Systems Research, 2001, 59, (2), pp. 121-129

46 Sang, Y., Shang, L., Wang, Z., Liu, C., and Yang, M.: 'Bayesian-combined wavelet regressive modeling for hydrologic time series forecasting', Chinese Science Bulletin, 2013, 58, (31), pp. 3796-3805