Feature Extraction Via Multiresolution MODWT Analysis in a Rainfall Forecast System

Fulgencio Buendia*†, A.M.Tarquis‡, G. Buendía § and D. Andina†
* GMV Defence & Security.
Madrid, Spain.
Email: fbuebue@yahoo.es, fsbb@gmv.es
†Universidad Politécnica de Madrid
Departamento de Señales, Sistemas y Radiocomunicaciones, E.T.S.I. Telecomunicación
Madrid, Spain
Email: diego@gc.ssr.upm.es
‡Department Matemática Aplicada
ETSI Agronomos
UPM, Spain
§ AEMET (Agencia Española de Meteorología)
Valladolid, Spain
Email:g.buendia@inm.es

Abstract— **During 30 years, expert meteorologists have been sampling meteorological measurements directly related to the rainfall event, in order to improve the current forecast procedures. This study performs the Feature Extraction and Feature Selection processes to extract the relevant information in the rainfall event. The Feature Extraction has been performed with a Multiresolution Analysis applying the Maxima Overlap Wavelet Transform. The selection of the wavelet decomposition, was obtained applying a Sequential Feature Selection algorithm based on General Regression Neural Networks. In this paper, it is also presented a novel architecture to perform short and medium term weather forecasts based on Neural Networks and time series estimation filters. The preliminary results obtained, present this architecture as a feasible alternative to the current forecast procedures performed by super computer simulation centers.**

I. Introduction

The Rainfall is one of the most important events in daily life of human beings, conditioning most of the activities either in the countryside or in big cities. Traffic, floods, electric power consumption, quality of the air, performance of public transportation are only some examples where the rainfall event has a direct impact.

During several decades, scientists have been trying to characterize the weather. Current forecasts are based on high complex dynamic models, see [1]-[7], that try to predict the atmospheric evolution based on a starting situation performing numerical simulations. The numerical calculations are processed by supercomputers, as an example, figure 1, shows the architecture of the European Center for Medium-range Weather Forecasts (ECMWF) (this center, placed in the UK, is supported by 31 countries).

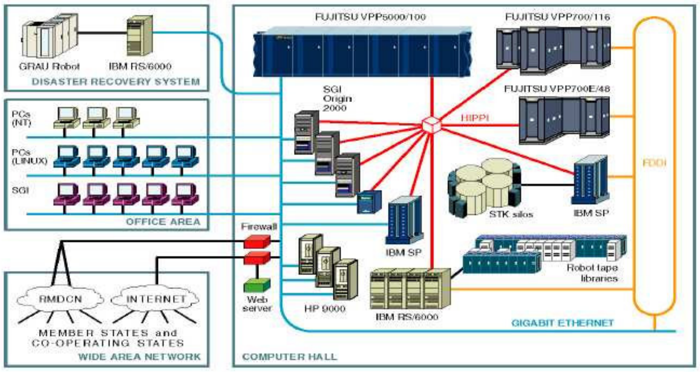

Fig. 1. European Centre for Medium-range Weather Forecast Architecture

As the weather evolution highly depends on the initial conditions ([5],[6]), even the minimum deviation either in the measurements (to set the initial state) or in the model definition shall produce a divergence between the forecast and the real evolution of the system. Actually, those are the sources of forecast errors in the current predictions, the initial state and model uncertainties.

Other approximations are [8]-[13] based on stochastic modelling and recently some attempts based on neural networks [14]-[15].

As a new approximation to the problem, in this paper, a novel system architecture is presented, as well as the early results obtained in the Feature Selection process. The goal of the study is to design and implement a short and medium term forecast system based on neural networks trained with the meteorological observations obtained in the meteorological

[1]_ _ _ _ _ _ _ _ __

This research has been partially supported by the National Spanish Institution "Ministerio de Educación Y Ciencia" as part of the project AGL2006-12689/AGR

Special Thanks to Patricia Tabares (University of Valladolid) for her corrections and kind comments

observatory of Valladolid (Spain), see figure 2.

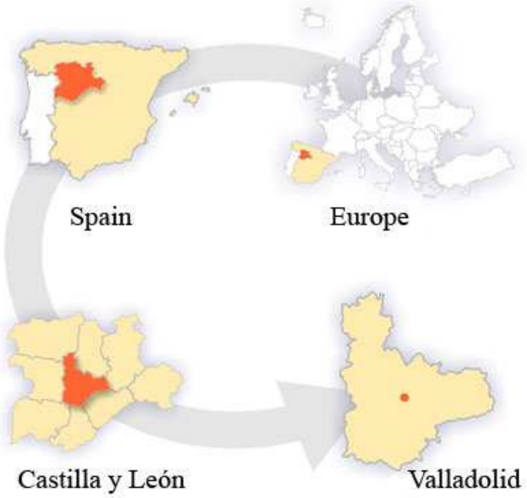

Fig. 2. Location of Valladolid

This paper describes the sequential feature selection procedure (SFS) that is being applied to the observations, in order to assess:

- That the observed variables contain all the necessary information to detect the rainfall events.
- The relevant information to estimate the rainfall event and intensity. This have been obtained by a Multiresolution Analysis (MRA) decomposition.

The paper is divided into different sections. Chapter two introduces the measures obtained and why they are relevant. Chapter three outlines the design of the rainfall forecaster. Chapter four contains a brief introduction to the MODWT MRA analysis. Chapter five describes the Feature Selection process. In chapter six, some conclusions are outlined as well as the future lines in the project.

II. Measurements Selection and Data Acquisition

As it has been introduced, three variables were selected to detect the different kinds of rainfall events, synoptic and convective rainfall events, for a classification of the different rainfall events please refer to [1]. Convective rainfalls are related to the evaporation processes; synoptic rainfall event are the result of cold fronts movements.

The selected variables are:

- Pressure waves (P_{ESC}) (measured in millibars or Hecto Pascals, Hp).
- Geopotential height at 500 Hp (Z_{500})(measured in meters).
- Thickness from 500 to 1000 Hp waves (H_{ESC})(measured in meters).

These measurements, suggested by the meteorologists, are supposed to contain enough information to forecast the rainfall events since they give information about pressure waves formation, neighborhood and pass through of warm and cold fronts, etc. Actually, the first aim of this study was to study if the variables were properly selected. The variables have been captured twice a day during 30 years. In picture 3, the time series of the measurements are represented as well as the rainfall intensity those days.

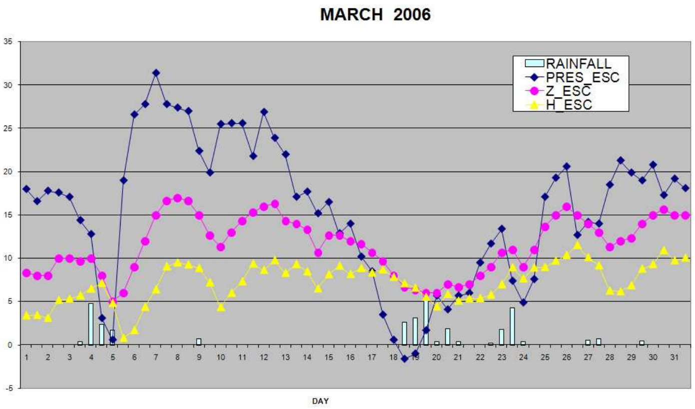

Fig. 3. Time series obtained in May 2006

P_{ESC} are obtained transforming the barometric measures reduced to sea level.

Z_{500} are obtained performing a linear extrapolation from the synoptic maps over the observatory of Valladolid. The synoptic maps represent the spatial distribution of multiple meteorological variables, measured at the same time all over the world in the meteorological stations. In this case, the maps represents the isobars lines (same pressure level curves) and geopotential height level curves. Figure 4 is an example of these maps.

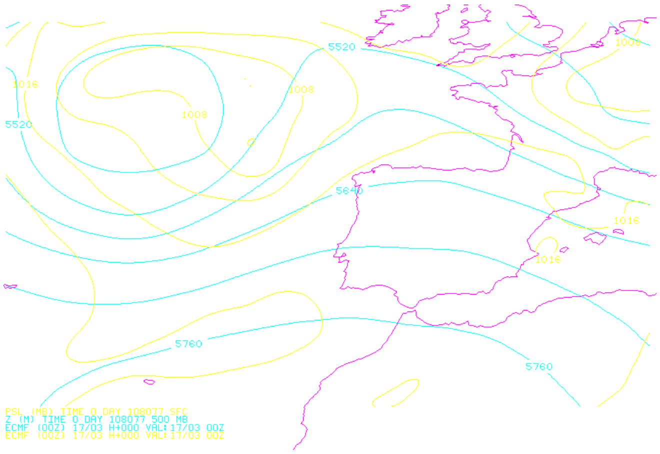

Fig. 4. Synoptic Map over Spain

H_{ESC} are obtained from the former measures with the following approximation (each Hp corresponds to 8.2 meters):

$$H_{ESC} = Z_{500} - (P_{ESC} - 1000) * 8.2 \quad (1)$$

This last variable is especially important since is directly related to the arrival of cold or warm fronts.

In addition to this measurements it has been sampled the wind direction in the observatory of Valladolid, but this data, is not been used currently.

III. System Design Overview

In this section, it is outlined an overview of the forecaster. The following figure shows the functional blocks that form the overall design:

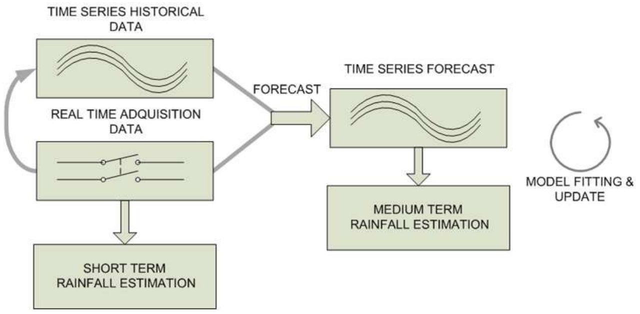
TIME SERIES HISTORICAL DATA
REAL TIME ADQUISITION DATA
FORECAST
TIME SERIES FORECAST
MODEL FITTING & UPDATE
MEDIUM TERM RAINFALL ESTIMATION
SHORT TERM RAINFALL ESTIMATION

Fig. 5. System Design Overview

As it can be seen, there is a Data Acquisition block, that shall automatically sample the data to perform three actions:

- Store the new observations into the database.
- Perform the Short Term rainfall estimation.
- Perform a forecast of the input variables, in order to estimate the rainfall events at medium term.

In the present paper, it is presented the feature extraction and the feature selection process, that are the first stages to design the rainfall estimator, see figure 6.

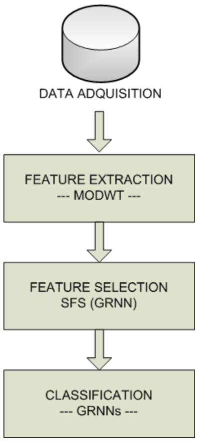

Fig. 6. Rainfall estimator design phases

Note that as not all the information contained in the time series is useful, it only will be necessary to forecast the time series with the relevant information.

IV. Maxima Overlap DWT Multiresolution Analysis

The Maxima Overlap Discrete Wavelet Transform (MODWT) ([16],[17],[18]), is a modified version of the Discrete Wavelet Transform (DWT) ([19]). However the DWT is orthogonal, the MODWT is highly redundant. As it is explained later, the MODWT has some quite appealing properties that makes it ideal to perform an analysis to a time series.

Both to DWT and MODWT, allows to perform a Multiresolution Analysis (MRA), that is a scale-based additive decomposition; it lets to decompose the time series into a sum of simpler time series, named Smooths S_j and Detail D_J, [16].Being $X(t)$ a time series, it can be written as:

$$X = \sum_{j=1}^{J-1} D_j + S_J \tag{2}$$

The name "Details and Smooths" came from the idea that D_j indicates the changes at scale $j-1$ and S_J contains the mean level at scale $J-1$.

The MODWT definition is obtained directly from the DWT: let be, $\tilde{h_{j,l}}$ the DWT wavelet filter and $\tilde{g_{j,l}}$ the scaling filter, being $l = 1..L$ the length of the filter and j the level of decomposition. The MODWT wavelet $h_{j,l}$ and scaling $g_{j,l}$ filters are directly defined:

$$\begin{aligned} h_{j,l} &= \tilde{h_l}/2^{j/2} \\ g_{j,l} &= \tilde{g_j}, l/2^{j/2} \end{aligned} \tag{3}$$

Then, the MODWT wavelet coefficients of level j are defined as the convolution of the time series and the MODWT filters:

$$\begin{aligned} W_{j,t} &= \textstyle\sum_{l=0}^{L-1} h_{j,l} X_{t-lmodN} \\ V_{j,t} &= \textstyle\sum_{l=0}^{L-1} g_{j,l} X_{t-lmodN} \end{aligned} \tag{4}$$

Note that from the above expressions, the MODWT wavelet coefficients at every scale shall have the same length as the original signal X. Now, (4) can be expressed in matrix notation as:

$$\begin{aligned} \vec{W_j} &= \bar{\bar{W}}_j \vec{X} \\ \vec{V_j} &= \bar{\bar{V}}_j \vec{X} \end{aligned} \tag{5}$$

Then, the original time series X can be expressed as (2) defining the Smooths and Details as follows:

$$\begin{aligned} D_j &= \bar{\bar{W}}_j^{\,T} \vec{W_j} \\ S_j &= \bar{\bar{W}}_j^{\,T} \vec{V_j} \end{aligned} \tag{6}$$

The DWT analysis of a time series depends critically from the starting sample, that means that the analysis of a time series and the same one starting one sample later is rather different ([16]). This is not true for the MODWT, actually the MODWT is also known as "shift invariant Wavelet Transform". So the the MRA obtained with the MODWT is "shift invariant". On the other hand, however the DWT is properly defined for sample sizes power of two, the MODWT is properly defined for any size samples. For the two above reasons, to perform the feature selection, a MRA MODWT analysis has been applied (instead of the DWT), to decompose P_{ESC}, Z_{500} & H_{ESC} into simpler components and extract the relevant information in the Rainfall event.

Figure 7 shows the MRA analysis to (P_{ESC}) with the Haar wavelet at level 3. At the left side of the figure, it is a label indicating that the first series is P_{ESC}, then the following four are the MRA Details and Smooths (S_1,S_2,S_3 & D_3) and the

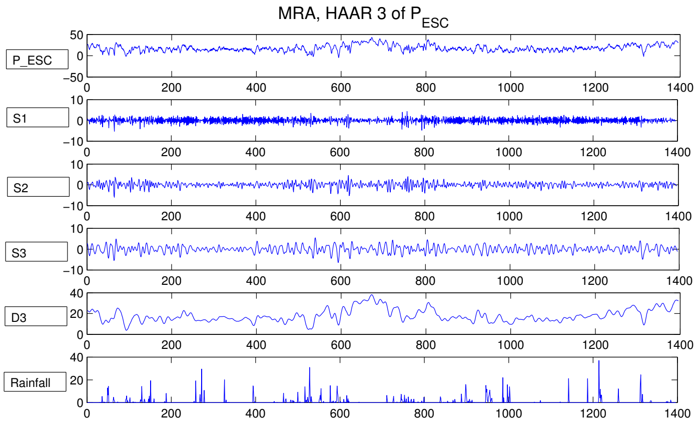

Fig. 7. MRA of the Pressure waves from February 2006 to December 2007

last one is the Rainfall intensity. The X axis is numbered form 1 to 1400, that corresponds to the samples from February 2006 to December 2007.

In order to know what is the better wavelet filter and level of decomposition to perform the MRA, it was performed a Sequential Feature Selection (SFS) algorithm, that is described in the following point.

V. SFS Process - Wavelet Filter Selection

To select the wavelet filter, and the level of decomposition, a SFS algorithm was implemented [20], selecting the filter and the decomposition level that obtained the better performance in unsupervised rainfall detection using a GRNN Network ([21],[22],[23]). GRNN structures, are a regression method proposed by Nadaraya-Watson and introduced by Specht in [21]. The principal advantages of GRNN are fast learning and convergence to the optimal regression surface as the number of samples becomes very large. The general expression of the GRNN network is:

$$\hat{y}(\mathbf{x}) = \frac{\sum_{q=1}^{M} y^{(q)} \exp(-\frac{D_q^2}{2\sigma_i^2})}{\sum_{q=1}^{M} \exp(-\frac{D_q^2}{2\sigma_i^2})} \qquad (7)$$

This kind of networks are quite simple to implement, caption 8 shows the generic architecture of this kind of networks, and are being used in time series forecasting [25],[24].

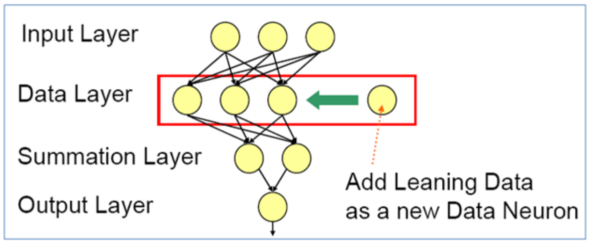

Fig. 8. GRNN Arquitecuture

The SFS was performed selecting among 20 different wavelet filters from level 2 to level 5 (for the haar wavelet, the time window of level 5 analysis would correspond to 16 days, 2^5 intervals of 12 hours). It was assumed that events two weeks ago are not relevant for daily rains, or at least will be better indicators.

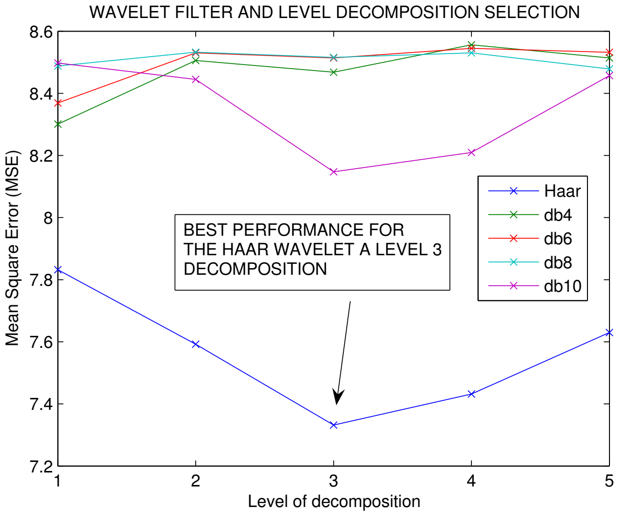

Fig. 9. Some SFS curves

To select the filter, the input to the GRNN was the wavelet decomposition by the different filters (Haar, Daubechies, etc) of P_{ESC}, Z_{500} and H_{ESC} at different scales; the signal to estimate, was the rainfall signal. The SFS was performed with the 2006 data, obtaining that the best transformation ,to estimate the rainfall signal, was the Haar one at level 3. Figure 9 shows the results obtained in the SFS, the Haar wavelet decomposition at level 3, achieves the best performance (minimum Mean Square Error (MSE)), the other wavelet filters (in the figure db4, db6,db8 and db10) present more MSE.

Once performed the feature selection, we tried to estimate the rainfall signal with the obtained decomposition; at first time, rainfall intensity estimation rates were not quite successful (too noisy), but after making binary the signal to estimate, ('1' -> rain event, '0' -> no rain), the detection of the rain event rate was about a 72% of success, both in 2006 and in 2007, (the feature selection was performed with the 2006 data, in order to test if the results were general).

These experiments suggest that the SFS process was quite successful to detect the rain event, so the variables are properly selected as it was suggested by the meteorologists. On the other hand, it has been outlined that it is necessary to divide the estimator into two stages, the first one detecting the rainfall event, and the second one estimating the intensity when necessary. As a third conclusion, as the more training data, the better GRNN performance presents, it shall be necessary to repeat the SFS process with a bigger time window, in order to accurate the detection percentage. Once fitted this stage of the overall forecast system, presented in section 3, it would be necessary to study the time series obtained in the MRA and selected in the SFS.

VI. Conclusions and Future Works

In this paper, it has been presented the SFS process performed to a bi-annual time series, in order to forecast rain events. The first results obtained, show that the selection of the measurements was right as well as it indicates that it is necessary to split the estimation network into two stages, the first one to detect the event and the second one to estimate the rainfall intensity. Another conclusion that can be deduced from the results is that the training set shall be larger, and not only based on an annual basis.

It has been presented the overall architecture of the rainfall forecast system. This system, that is in its first development stages, is designed to estimate the rainfall event at short and medium term. The promising results present this architecture based in Neural Networks, as a feasible alternative to the big supercomputers centers.

In addition, this novel approach does not have the inherent limitations of modelling the climate in supercomputers centers (uncertainties in the measurements and model definitions), so the accuracy of the final system will be only limited by the time window loaded in the database of the system.

References

[1] James R. Holton, "An Introduction To Dynamic Meteorology," Academic Press; 4rd edition, April 2004, ISBN: 978-0123540157

[2] Howard Bluestein, "Synoptic-Dynamic Meteorology in Midlatitudes: Principles of Kinematics and Dynamics," Oxford University Press; May 1992, ISBN: 978-0195062670

[3] Tim Vasquez, "Weather Forecasting Handbook," Weather Graphics Technologies; 5th edition, June 2002, ISBN: 978-0970684028

[4] McGuffie, Kendal, *A climate modelling primer*, 2005, ISBN: 978-0-470-85751-9, JOHN WILEY & SONS, LTD.

[5] Edward N Lorenz, "The nature and theory of the general circulation of the atmosphere," World Meteorological Organization, 1961

[6] Ales Raidl, "Is Weather Chaotic?,"*Perspective in Modern Prediction Methods*, 44-48, 1998, Technical University of Brno, ISBN: 80-214-1222-4

[7] L.A. López Álvarez, "La Predicción del Tiempo a Partir de los Modelos Numéricos," *: Revista de Meteorología de Colombia.* , vol. 6, pp $1 - 8$, ISSN 0124-6984.

[8] Cowpertwait P. S. P., OConnell P. E., Metcalfe A. V., and Mawdsley J. A., "Stochastic point process modelling of rainfall, I. Single site fitting and validation," *: Journal of Hydrology* , vol. 175, pp $17 - 46$, 1996.

[9] Rodriguez-Iturbe I., Cox D. R., and Isham V., a.D. Marr and T. Poggio, “Some models for rainfall based on stochastic point processes," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 410, No. 1839 (Apr. 8, 1987), pp. 269-288

[10] R. Coe and R. D. Stern,“Fitting Models to Daily Rainfall Data," *Journal of Applied Meteorology* , Volume 21, Issue 7 (July 1982), pp. 1024-1031

[11] Craig S. Thompsona, Peter J. Thomsonb and Xiaogu Zhenga, “Fitting a multisite daily rainfall model to New Zealand data,"*Journal of Hydrology*, Volume 340, Issues 1-2, 30 June 2007, pp. $25 - 39$.

[12] Bellie Sivakumar, “Fitting a multisite daily rainfall model to New Zealand data," *Hydrological Sciences-Journal-des Sciences Hydrologiques* 45(5) October 2000.

[13] T. Beer,“Modelling rainfall as a fractal process," *Mathematics and Computers in Simulation*, , 1989, 32(1/2), pp. $119 - 124$.

[14] Y. Quan, L. Yuchang , "Research on weather forecast based on neural networks,"*Intelligent Control and Automation, 2000. Proceedings of the 3rd World Congress on*, vol. 2, pp. 1069-1072, 2000, ISBN: 0-7803-5995-X.

[15] M. Miwa, S. Makoto, T. Eiichiro, MACKIN K J, "Construction of a Weather Forecast System using Neural Networks trained by Genetic Algorithm.," *Faji Shisutemu Shinpojiumu Koen Ronbunshu*, ISSN:1341-9080 VOL.16 ; pp.139-142, 2000

[16] Donald B. Percibal & Adrew Walden, *Wavelet Methods for Time Series Analysis*, 2002, Cambridge University Press.

[17] P. Crowley, M. Patrick, J. Lee, "Decomposing the Co-Movement of the Business Cycle: A Time-Frequency Analysis of Growth Cycles in the Euro Area", Bank of Finland Research Discussion Paper No. 12/2005

[18] A. T. Walden and A. Contreras Cristan, "Matching pursuit by undecimated discrete wavelet transform for non-stationary time series of arbitrary length.,"*Statistics and Computing*, Springer Netherlands, ISSN 0960-3174, VOL. 8, Number 3, 1998

[19] S. Mallat,*"A Wavelet Tour of Signal Processing"*, Academic Press, 2nd edition, 1999. ISBN-10: 012466606X

[20] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2001.

[21] Specht, D. A. (1991). “A General Regression Neural Network," *IEEE Transactions on Neural networks*, vol. 2, no. 6. pp. 568-576, 1991.

[22] F. Buendia Buendia, M. Barrón-Adame Antonio Vega-Corona and Diego Andina Título, "Improving GRNNs in CAD Systems", *Lecture Notes in Computer Science, ISBN 978-3-540-23056-4, pp. 160-167, 2004*

[23] Leung, Mark.T.; Chen, An-Sing; Daouk, Hazem, “Forecasting exchange rates using general regression neural networks," Computers & Operations Research,, vol. 27, pp. $1093 - 1110$*, 2000.*

[24] Weimin Li, Jianwei Liu and Jiajin Le, "Using GARCH-GRNN Model to Forecast Financial Time Series", Lecture Notes in Computer Science, ISBN 978-3-540-29414-6, pp. 565-574, 2005

[25] Weimin Li, Yishu Luo, Qin Zhu, Jianwei Liu and Jiajin Le, "Applications of AR-GRNN model for financial time series forecasting", Neural Computing & Applications, Springer London, Augost 2007 (published online)*