Wavelet Autoregressive Model for Monthly Sardines Catches Forecasting Off Central Southern Chile

Nibaldo Rodriguez, Jose Rubio, and Eleuterio Yañez

Pontificia Universidad Católica de Valparaíso, Chile
nibaldo.rodriguez@ucv.cl

Abstract. In this paper, we use multi-scale stationary wavelet decomposition technique combined with a linear autoregressive model for one-month-ahead monthly sardine catches forecasting off central southern Chile.The monthly sardine catches data were collected from the database of the National Marine Fisheries Service for the period between 1 January 1964 and 30 December 2008. The proposed forecasting strategy is to decompose the raw sardine catches data set into trend component and residual component by using multi-scale stationary wavelet transform. In wavelet domain, both the trend component and the residual component are independently predicted using a linear autoregressive model. Hence, proposed forecaster is the co-addition of two predicted components. We find that the proposed forecasting method achieves a 99% of the explained variance with a reduced parsimonious and high accuracy.

Keywords: forecasting, wavelet decomposition, autoregression.

1 Introduction

Common sardine is an important fish resource for industrial in the central southern area off Chile. In order to develop sustainable exploitation policies, forecasting the stock and catches of sardines in Chile is one of the main goals of the fishery industry and the government. However, fluctuations in the environmental variables complicate this task. To the best of our knowledge, few publications exist on forecasting models for pelagic species. In recent years, linear regression models [1,2] and artificial neuronal networks (ANN) [3,4] have been proposed for forecasting models. The disadvantage of models based on linear regressions is the supposition of stationarity and linearity of the time series of pelagic species catches. Although ANN allow modeling the non-linear behavior of a time series, they also have some disadvantages such as slow convergence speed and the stagnancy of local minima due to the steepest descent learning method. To improve the convergence speed and forecasting precision of anchovy catches off northern Chile, Gutierrez [3] proposed a hybrid model based on a multilayer perceptron (MLP) combined with an autoregressive integrated moving average model. The architecture of the MLP consists of an input layer with 6 nodes, two hidden layers

C. San Martin and S.-W. Kim (Eds.): CIARP 2011, LNCS 7042, pp. 654–663, 2011.

of 15 nodes each, and an output layer with one node; the Levenberg Maquardt (LM) method was used as the learning method. This forecaster obtained a coefficient of determination R^2 of 82%, which improved slightly when combining the MLP model with the ARIMA model, reaching an R^2 of 87%. One of the disadvantages of this hybrid model is its high parsimony (230 parameters) and low forecasting precision. In this paper, the proposed forecasting model is based on multi-scale wavelet decomposition combined with autoregressive models. The multi-scale wavelet decomposition technique was selected due to its popularity in hydrological [5,6], electricity market [7], financial market [8] and smoothing methods [9,10,11]. This wavelet technique is based on the discreet wavelet transform (DWT) or the stationary wavelet transform (SWT) [12]. The advantage of these wavelet transforms in non-stationary time series analysis is their capacity to separate low frequency (LF) from high frequency (HF) components. Whereas the LF component reveals long-term trends, the HF component describes short-term fluctuations in the time series. Being able to separate these components is a key advantage in proposed forecasting strategies since the behavior of each frequency component is more regular than the raw time series.

Therefore, an one-month-ahead monthly sardines catches forecasting scheme is proposed. The forecasting strategy is to decompose the raw sardine catches data set into trend component and residual component by using multi-scale stationary wavelet transform (SWT). In wavelet domain, both the trend component and residual component are independently predicted using a linear autoregressive model.

This paper is organized as follows. In the next section, we briefly describe the multi-scale stationary wavelet transform and the proposed multi-scale wavelet autoregressive forecasting model. The simulation results and performance evaluation are presented in Section 3 followed by conclusions in Section 4.

2 Proposed Forecasting Model

This section presents the proposed forecasting model for one-month-ahead sardines catches off central-southern Chile. Moreover, instead of using the raw data set of past observations to predict the future value $x(n+1)$, we use its wavelet coefficients.

2.1 Stationary Wavelet Decomposition

A signal $x(n)$ can be represented at multiple resolutions by decomposing the signal on a family of wavelets and scaling functions [9,10,11]. The approximation (scaled) signals are computed by projecting the original signal on a set of orthogonal scaling functions of the form:

$$\phi_{jk}(t) = \sqrt{2^{-j}}\phi(2^{-j}t - k) \tag{1}$$

or equivalently by filtering the signal using a low pass filter of length r, $h = [h_1, h_2, ..., h_r]$, derived from the scaling functions. On the other hand, the detail signals are computed by projecting the signal on a set of wavelet basis functions of the form

$$\psi_{jk}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - k) \tag{2}$$

or equivalently by filtering the signal using a high pass filter of length r, $g = [g_1, g_2, ..., g_r]$, derived from the wavelet basis functions. Finally, repeating the decomposing process on any scale J, the original signal can be represented as the sum of all detail coefficients and the last approximation coefficient.

In time series analysis, discrete wavelet transform (DWT) often suffers from a lack of translation invariance. This problem can be tackled by means of the un-decimated stationary wavelet transform (SWT). The SWT is similar to the DWT in that the high-pass and low-pass filters are applied to the input signal at each level, but the output signal is never decimated. Instead, the filters are up-sampled at each level.

Consider the following discrete signal $x(n)$ of length N where $N = 2^J$ for some integer J. At the first level of SWT, the input signal $x(n)$ is convolved with the $h_1(n)$ filter to obtain the approximation coefficients $a_1(n)$ and with the $g_1(n)$ filter to obtain the detail coefficients $d_1(n)$, so that:

$$a_1(n) = \sum_k h_1(n-k)x(k) \tag{3a}$$

$$d_1(n) = \sum_k g_1(n-k)x(k) \tag{3b}$$

because no sub-sampling is performed, $a_1(n)$ and $d_1(n)$ are of length N instead of $N/2$ as in the DWT case. At the next level of the SWT, $a_1(n)$ is split into two parts by using the same scheme, but with modified filters h_2 and g_2 obtained by dyadically up-sampling h_1 and g_1.

The general process of the SWT is continued recursively for $j = 1, ..., J$ and is given as:

$$a_{j+1}(n) = \sum_k h_{j+1}(n-k)a_j(k) \tag{4a}$$

$$d_{j+1}(n) = \sum_k g_{j+1}(n-k)a_j(k) \tag{4b}$$

where h_{j+1} and g_{j+1} are obtained by the up-sampling operator inserts a zero between every adjacent pair of elements of h_j and g_j; respectively.

Therefore, the output of the SWT is then the approximation coefficients a_J and the detail coefficients $d_1, d_2, ..., d_J$, whereas the original signal $x(n)$ is represented as a superposition of the form:

$$x(n) = a_J(n) + \sum_{j=1}^{J} d_j(n) \tag{5}$$

The wavelet decomposition method is fully defined by the choice of a pair of low and high pass filters and the number of decomposition steps J. Hence, in this study we choose a pair of haar wavelet filters [12]:

$$h = \left[\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}\right] \tag{6a}$$

$$g = \left[\frac{-1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}\right] \tag{6b}$$

On the other hand, a key issue for the success of any wavelet forecasting model is suitable selection of the J level decomposition. At higher J, the variability of a large number of predicted data is lower, so their prediction is easier and accurate. In our proposed model, we determine the value of J using a stopping criterion that is given as:

$$\rho = \frac{Pd_j}{P_x} < \epsilon \tag{7}$$

where Pd_j and P_x represents the average power of the detail component $d_j(n)$ and the original data $x(n)$, respectively.

We stop the decomposition on the level for which the ρ ratio is substantially less than a threshold ϵ. The choice of the value of ϵ is not clear from a physical point of view and different sets of approximation coefficients will be produced by the wavelet decomposition method for different values of ϵ. In order to obtain accurate and parsimonious forecasting results, the value of ϵ was set to 0.0 in this work.

Finally, wavelet scales are such that times are separated by multiples of $2^j, j = 1, ..., J$. Our data set involves monthly observations so that the wavelet scales are such that scale 1 is associated with $1-2$ month dynamics, scale 2 with $2-4$ month dynamics, scale 3 with $4-8$ month dynamics, scale 4 with $8-16$ month dynamics, and so on.

2.2 Wavelet Autoregressive Model

In order to predict the future signal $x(n+1)$, we can separate the original signal $x(n)$ into two components. The first component presents the trend $t(n)$ of the series and is characterized by slow dynamics, whereas the second component presents the residue $r(n)$ of the series and is characterized by fast dynamics. Therefore, our forecasting model will be the co-addition of two predicted values given as:

$$x(n+1) = t(n+1) + r(n+1) \tag{8}$$

On the one hand, the residual component is estimated using a linear autoregressive (AR) model given as:

$$r(n+1) = \sum_{j=1}^{J} \sum_{i=1}^{m} \alpha_{ji} d_j[n-i+1] \tag{9a}$$

where the J value denotes the level of stationary wavelet decomposition and the m value represents the autoregressive order of the detail coefficients.

On the other hand, the trend component is estimated using a linear AR model given as:

$$t(n+1) = \sum_{i=1}^{m} \beta_i a_J[n-i+1] \tag{10}$$

We propose estimating the linear parameters $\theta = \{\alpha_i, \beta_i\}$ using the least squares method based on the Moore-Penrose pseudo-inverse. If we suppose a set of N_s training input-output samples, then we can perform N_s equations of the form of (9) and (10) as follows:

$$\Re = \alpha \Phi \tag{11a}$$

$$\Gamma = \beta \Psi \tag{11b}$$

where

$$\Phi = [d_1(n), \cdots, d_1(n-m+1), \cdots, d_J(n-m+1)] \tag{12a}$$

$$\Psi = [a_J(n), a_J(n-1), \cdots, a_J(n-m+1)] \tag{12b}$$

The optimal values of the linear parameters α_i and β_i are obtained using the following residual sum of squares (RSS) function defined as:

$$RSS(\alpha) = \sum_{n=1}^{N_s} \left[R(n+1) - r(n+1)\right]^2 \tag{13a}$$

$$R(n+1) = x(n+1) - a_J(n+1) \tag{13b}$$

$$RSS(\beta) = \sum_{n=1}^{N_s} \left[a_J(n+1) - t(n+1)\right]^2 \tag{13c}$$

The result of minimizing the RSS objective function is:

$$\alpha = (\Phi^T \Phi)^{\dagger} \Phi^T \Re \tag{14a}$$

$$\beta = (\Psi^T \Psi)^{\dagger} \Psi^T \Gamma \tag{14b}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse [13].

Once we have decided upon a forecasting structure to use, the next task is to determine the autoregressive order on the different scales. This can be done using the criterion given by the ration of the mean absolute deviation to the mean of the time series (MADM) versus lagged values of the predictor variables, where the MADM value is defined as [14]:

$$MADM = \frac{\sum_{i=1}^{N_s} |A_i - F_i|}{\bar{A}} \tag{15}$$

where A_i is the actual value at time i , F_i is the forecasted value at time i, $\bar{A}$ is the mean value of observed monthly catches, and N_s is the number of samples.

3 Experiments and Results

In this section, we apply the proposed strategy for 1-month-ahead forecasting of the monthly catches of sardines. The data set used corresponded to sardine landings off central southern area Chile. These samples were collected monthly from 1 January 1964 to 30 December 2008 by the National Fishery Service of Chile (www.sernapesca.cl).

The proposed linear wavelet autoregressive (WAR) forecasting model basically involves three stages. In the first stage, the original data set is decomposed into different wavelet scales by using stopping criterion given in (7) to separate both the trend component (approximation component) and the residual component (difference between original data and trend component). In the second stage, the trend component and residual component are independently forecasted by using an linear autoregressive model. In the third stage, the next sample is predicted by the co-addition of two predicted components.

The raw sardines data set have been normalized to the range from 0 to 1 by simply dividing the real value by the maximum of the appropriate set. On the other hand, the original data set was also divided into two subsets as shown in Fig.1. In the first subset, the data from 1 January 1964 to 30 December 2003 were chosen for the training phase ($N_s = 480$ months), whereas the remaining data

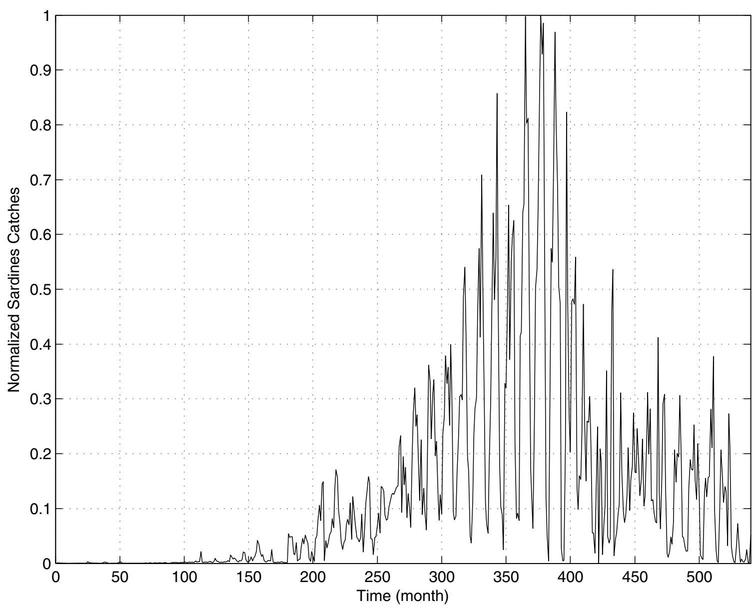

Fig. 1. Observed monthly sardine catches data from 1964 to 2008

were used for the validation phase. This normalized data set, when subjected to the stopping criterion (7), yielded a 5 level wavelet decomposition. The low frequency component a_5 represents the trend of the observed sardines catches data set. On the other hand, detail components $\{d_1, d_2, d_3, d_4, d_5\}$ contain high frequency components of the original data such that d_1 the highest frequency component and d_1 is considered to be more related to the noisy part of the observed data, whereas d_5 contains lower frequency information than $\{d_4, d_3, d_2\}$.

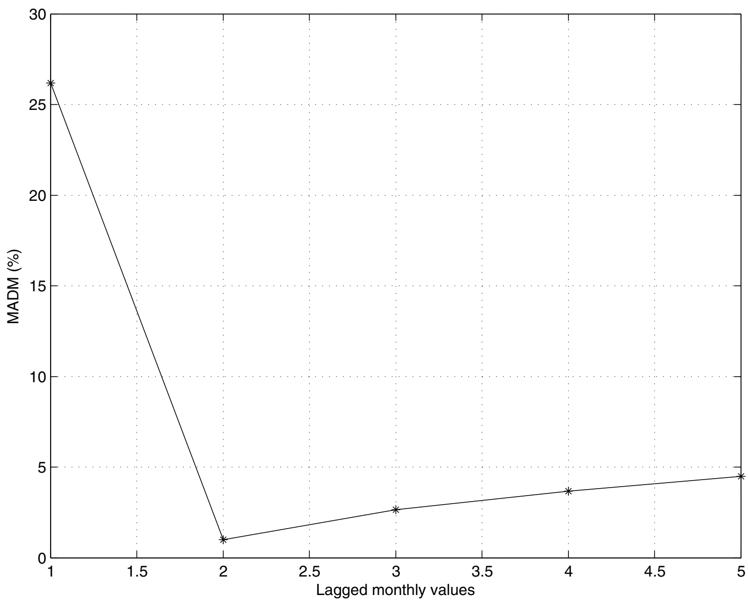

Fig. 2. Model order selection criteria

Hence, residual component forecasting is based on a linear AR model, whereas predicting the trend component also is done with a linear AR model. Once we chose the multi-scale autoregressive forecasting structure to use, the next task was to determine the autoregressive order by using the method given in (14) and (15). After we applied the least squares method and the MADM, we decided to use two lagged values on each level wavelet decomposition due to the parsimony principle and precision of the proposed WAR(J,m) model with $J = 5$ and $m = 2$ as shown in Fig.2. In this study, two criteria of forecasting accuracy were used to evaluate the forecasting capabilities of the WAR model. The first measurement is the coefficient determination (R^2) given as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N_s}(A_i - F_i)^2}{\sum_{i=1}^{N_s}(A_i - \bar{A})^2} \tag{16}$$

where A_i is the actual value at time i, F_i is the forecasted value at time i, $\bar{A}$ is the mean value of observed monthly catches, and N_s is the number of forecasts.

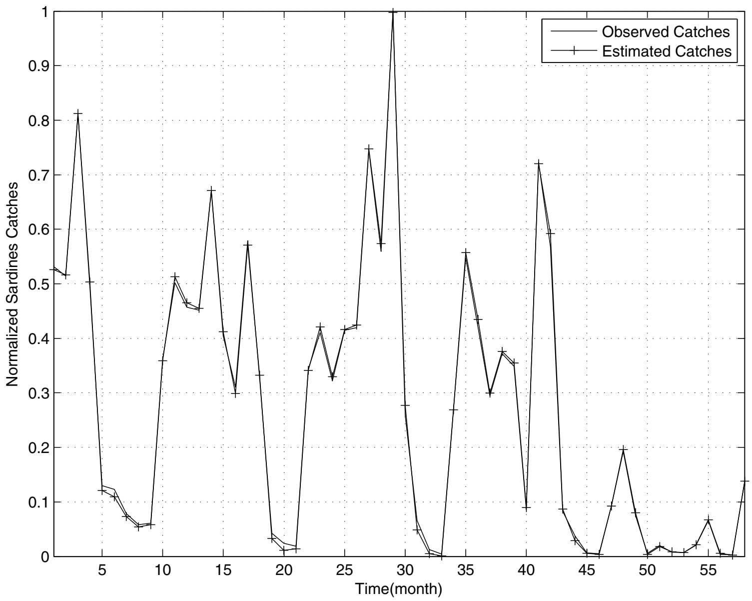

Fig. 3. Observed sardine catches vs estimated sardine catches from 2004 to 2008

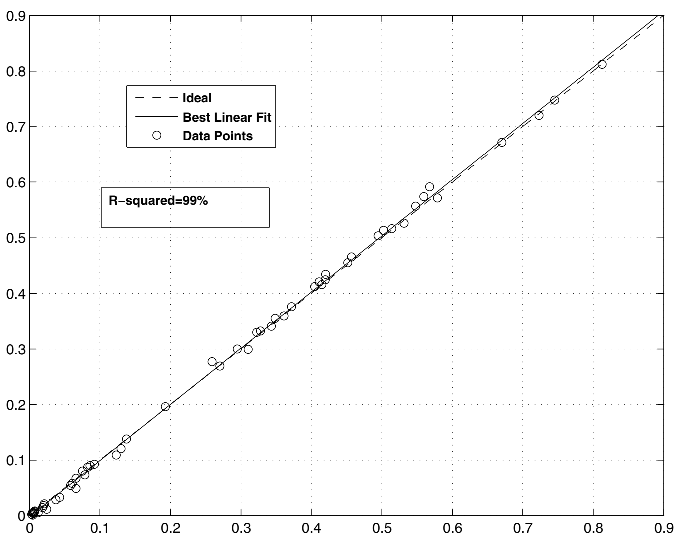

Fig. 4. Scatter for estimates monthly sardines catches

If R-square is large, then the model is good. Conversely, if R-square is small, then the model is bad.

The second criterion is the mean absolute percentage error (MAPE) given as:

$$MAPE(\%) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left| \frac{A_i - F_i}{F_i} \right| \times 100 \tag{17}$$

Figures 3 and 4 show the results obtained with the best WAR(5,2) forecasting model during the testing phase. Fig. 3 provides data on observed monthly sardine catches versus forecasted catches; this forecasting behavior is very accurate for testing data with a MAPE below 9.4%. Fig. 4 shows the regression between observed and estimated monthly sardine catches. The good fit of the data to line 1 : 1 and 99% of the explained variance can be seen in Fig.4. This level of explained variance was achieved due to use of multi-scale stationary wavelet decomposition.

4 Conclusions

In this paper was proposed a one-month-ahead monthly sardine catches forecasting strategy to improve prediction accuracy. The reason of the improvement in forecasting accuracy was due to use stationary haar wavelet decomposition to separate both the trend and residual components of the raw time series, since the behavior of each component is more smoothing than raw data set. It was show that the proposed forecaster achieves a MAPE value of 9.4% and a R-squared of 99%. Besides, proposed forecasting results showed that the 32 previous months contain valuable information to explicate a highest variance level for sardines catches forecasting. These months can be related with ocean-atmospheric aspects, which have a great influence on pelagic fish fisheries in Chile. Finally, wavelet-autoregressive forecasting strategy can be suitable as a very promising methodology to any other pelagic specie.

References

1. Stergiou, K.I.: Prediction of the Mullidae fishery in the easterm Mediterranean 24 months in advance. Fish. Res. 9, 67–74 (1996)
2. Stergiou, K.I., Christou, E.D.: Modelling and forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series methods. Fish. Res. 25, 105–138 (1996)
3. Gutierrez, J.C., Silva, C., Yaez, E., Rodriguez, N., Pulido, I.: Monthly catch forecasting of anchovy engraulis ringens in the north area of Chile: Nonlinear univariate approach. Fisheries Research 86, 188–200 (2007)
4. Garcia, S.P., DeLancey, L.B., Almeida, J.S., Chapman, R.W.: Ecoforecasting in real time for commercial fisheries:the Atlantic white shrimp as a case study. Marine Biology 152, 15–24 (2007)
5. Adamowski, J.F.: Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. Journal of Hydrology 353(3-4), 247–266 (2008)

6. Kisi, O.: Stream flow forecasting using neuro-wavelet technique. Hydrological Processes 22(20), 4142–4152 (2008)
7. Amjady, N., Keyniaa, F.: Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method. International Journal of Electrical Power Energy Systems 30, 533–546 (2008)
8. Bai-Ling, Z., Richard, C., Marwan, A.J., Dominik, D., Barry, F.: Multiresolution Forecasting for Futures Trading Using Wavelet Decompositions. IEEE Trans. on Neural Networks 12(4) (2001)
9. Coifman, R.R., Donoho, D.L.: Translation-invariant denoising, Wavelets and Statistics. Springer Lecture Notes in Statistics, vol. 103, pp. 125–150. Springer, Heidelberg (1995)
10. Nason, G., Silverman, B.: The stationary wavelet transform and some statistical applications, Wavelets and Statistics. Springer Lecture Notes in Statistics, vol. 103, pp. 281–300. Springer, Heidelberg (1995)
11. Pesquet, J.-C., Krim, H., Carfantan, H.: Time-invariant orthonormal wavelet representations. IEEE Trans. on Signal Processing 44(8), 1964–1970 (1996)
12. Percival, D.B., Walden, A.T.: Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge (2000)
13. Serre, D.: Matrices: Theory and applications. Springer, New York (2002)
14. Kolassa, S., W, S.: Advantages of the mad/mean ratio over the mape. The International Journal of Applied Forecasting (6), 40–43 (2007)