Multiscale RBF Neural Network for Forecasting of Monthly Hake Catches off Southern Chile

Nibaldo Rodriguez, Lida Barba and Jose Miguel Rubio L.

Abstract—We present a forecasting strategy based on stationary wavelet transform combined with radial basis function (RBF) neural network to improve the accuracy of 3-month-ahead hake catches forecasting of the fisheries industry in the central southern Chile. The general idea of the proposed forecasting model is to decompose the raw data set into an annual cycle component and an inter-annual component by using 3-levels stationary wavelet decomposition. The components are independently predicted using an autoregressive RBF neural network model. The RBF neural network model is composed of linear and nonlinear weights, which are estimates using the separable nonlinear least squares method. Consequently, the proposed forecaster is the co-addition of two predicted components. We demonstrate the utility of the proposed model on hake catches data set for monthly periods from 1963 to 2008. Experimental results on hake catches data show that the autoregressive RBF neural network model is effective for 3-month-ahead forecasting.

Index Terms—Neural network, forecasting, nonlinear least squares.

I. Introduction

THE highly productive coastal upwelling zone off central / southern Chile (30–40ºS) sustains a strong fishery based on hake catches. The hake is highly important for economic development in the southern zone of Chile. One of the main goals of the fishery industry and the governments is to develop sustainable exploitation policies. However, fluctuations in the marine ecosystem complicate this task. To the best of our knowledge, few publications exist on multi-step-ahead forecasting models for fisheries resources. In recent years, linear regression models [1], [2] and artificial neuronal networks (ANN) [3]–[5] have been proposed for fisheries 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 allows modeling the non-linear behaviour of a time series, they also have some disadvantages such as the stagnancy of local minimum due to the steepest descent learning method and over-fitting problem. A multilayer

Nibaldo Rodriguez is with the School of Computer Engineering at the Pontificia Universidad Católica de Valparaíso, Av. Brasil 2241, Chile (e-mail: nibaldo.rodriguez@ucv.cl).

Lida Barba is with the School of Computer Engineering at the Universidad Nacional de Chimborazo, Av. Antonio Jose de Sucre, Km 1.5, Ecuador.

Jose Miguel Rubio L. is with the School of Computer Engineering at the Pontificia Universidad Católica de Valparaíso, Av. Brasil 2241, Chile.

perceptron neural network to improve the convergence speed and forecasting accuracy of anchovy and sardines catches off northern Chile was proposed by [3]–[5], which reported a coefficient of determination between 82% and 87%.

In this paper, the stationary wavelet transform (SWT) combined with autoregressive RBF neural network models are applied to improve accuracy forecasting of monthly hake catches of the southern zone of Chile. The advantage of the SWT in non-stationary time series analysis is their capacity to separate low frequency (LF) from high frequency (HF) components [6]–[12]. Whereas the LF component reveals long-term trends, the HF component describes short-term fluctuations in the time series. To separate these components is a key advantage in the proposed forecasting strategies, since the behaviour of each frequency component is more regular than the raw time series.

On the other hand, neural network techniques based on the nonlinear least square method have been accepted in several domains as a flexible modeling technique, suited for capturing nonlinear relationships between predictor variables and a response variable. Therefore, the proposed forecaster decomposes the raw data set into annual cycle component and interannual component, which are predicted independently using a RBF neural network model, whereas the final forecasting results are the sum of results obtained from a single model.

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

II. Neural Forecasting Model Based On SWT

This section presents the proposed forecasting model for monthly hake catches off southern Chile. The proposed forecaster basically involves three stages. In the first stage, the original data set is decomposed into 3-level stationary wavelet decomposition to separate the annual cycle component and the interannual component. In the second stage, the annual cycle component and the interannual component are forecasted independently using a RBF neural network model with the separable nonlinear least squares (SNLS) algorithm. In the third stage, the future values are predicted by the co-addition of two predicted components.

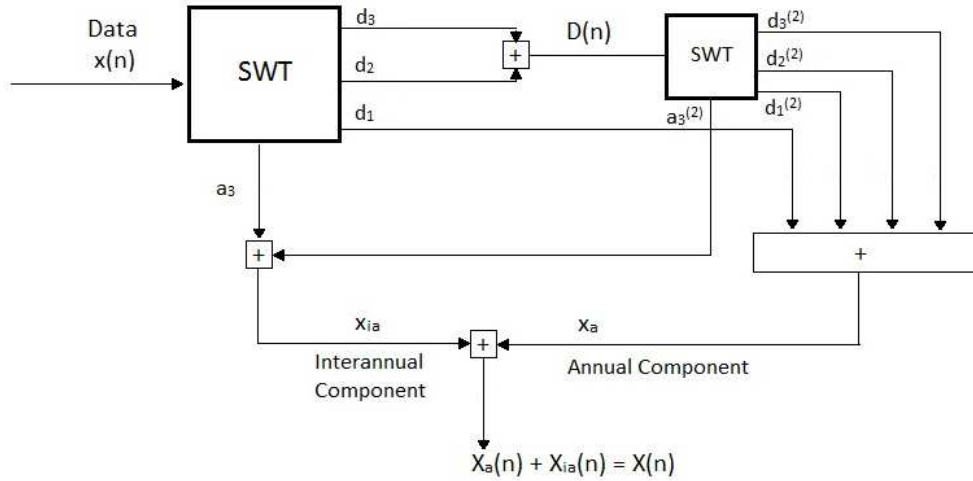Nibaldo Rodriguez, Lida Barba, Jose Miguel Rubio L.

Fig. 1. Extraction process of components

A. Stationary wavelet decomposition

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 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.

A signal can be represented at multiple resolutions by decomposing the signal on a family of wavelets and scaling functions [8]–[12]. The approximation (scaled) signals are computed by filtering the signal using a low pass filter of length r, $h = [h_1, h_2, ..., h_r]$. On the other hand, the detail signals are computed by filtering the signal using a high pass filter of length r, $g = [g_1, g_2, ..., g_r]$. 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.

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:

$$a_1(n) = \sum_k h_1(n-k)x(k) \tag{1a}$$

$$d_1(n) = \sum_k g_1(n-k)x(k) \tag{1b}$$

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. Here the up-sampling

operator inserts a zero between every adjacent pair of elements of $a_j(n)$.

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

$$a_{j+1}(n) = \sum_k h_{j+1}(n-k)a_j(k) \tag{2a}$$

$$d_{j+1}(n) = \sum_k g_{j+1}(n-k)a_j(k) \tag{2b}$$

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:

$$x(n) = a_J(n) + \sum_{j=1}^{J} d_j(n) \tag{3}$$

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 given as:

$$h = \left[\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}} \right] \tag{4a}$$

$$g = \left[\frac{-1}{\sqrt{2}} \ \frac{1}{\sqrt{2}} \right] \tag{4b}$$

B. Proposed forecasting model

In order to predict the future signal $\hat{x}(n + h)$, our direct forecasting model will be the co-addition of two predicted values given as:

$$\hat{x}(n+h) = \hat{x}_a(n+h) + \hat{x}_{ia}(n+h) + e(n) \tag{5}$$

where $h = 3$ represents the forecasting horizon, $\hat{x}_a$ represents the annual cycle component, $\hat{x}_{ia}$ denotes the inter-annual
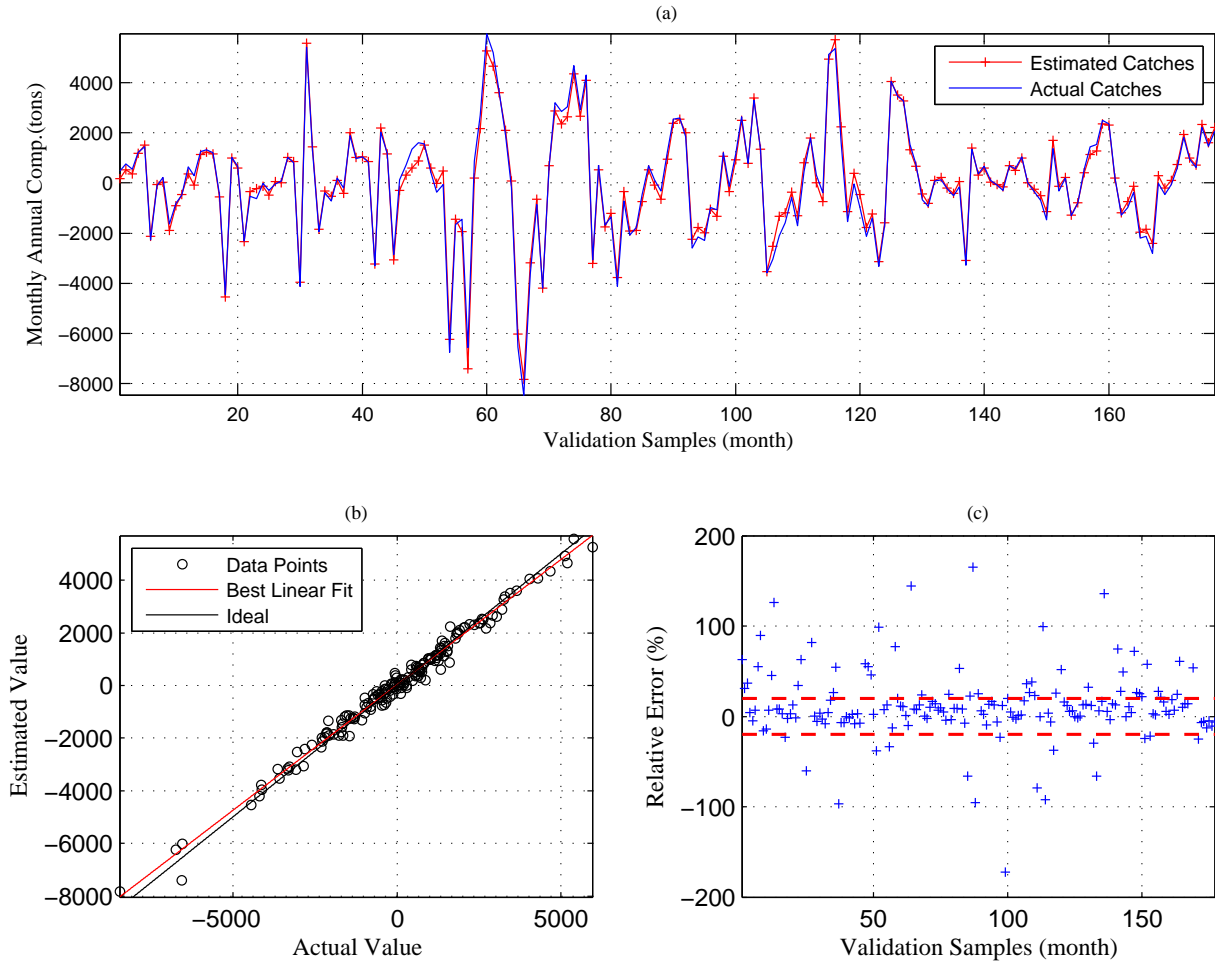
Fig. 2. Three-step-ahead forecasting for annual cycle component

component and e is an random process with distribution $\aleph(0, \sigma_e^2)$.

The components extraction process is described as follows and the flowchart is shows in Figure 1. The raw catches time series is decomposed using 3-level wavelet decomposition. The first detail component $d_1(n)$ of the direct output of the decomposition represents a (2-4)-month variability, the second detail component $d_2(n)$ of the direct output denotes a (4-8)-month dynamic, whereas the third detail component $d_3(n)$ contains some annual cycle as well as some large timescale variation with a (8-16) months dynamic. Therefore to extract annual cycle component $x_a(n)$, the detail components two and three of the direct output are combined $D = d_2 + d_3$ and then the new component D is subjected to 3-level wavelet decomposition.

The three detail components $d_1^{(2)}, d_2^{(2)}, d_3^{(2)}$ resulting from this wavelet decomposition are combined with the first detail component of the direct output d_1, whereas the third approximation component $a_3^{(2)}$ is combined with the third approximation component a_3 of the direct output to obtain the new inter-annual component x_{ia}.

The annual cycle component and the inter-annual component are estimated using a nonlinear autoregressive model with exogenous inputs given by the following equations

$$\hat{x}_a(n + h) = f([y_a(n), u_a(n)]) \qquad (6)$$

where $y_a(n) = [x_a(n), x_a(n - 1), ..., x_a(n - m)]$ are the endogenous inputs, $u_a(n) = [x_{ia}(n), x_{ia}(n - 1), ..., x_{ia}(n - m)]$ are the exogenous inputs, m represents the memory of the model and $f(\cdot)$ is a nonlinear function.

$$\hat{x}_{ia}(n + h) = g([y_{ia}(n), u_{ia}(n)]) \qquad (7)$$

where $y_{ia}(n) = [x_{ia}(n), x_{ia}(n - 1), ..., x_{ia}(n - m)]$ are the endogenous inputs, $u_{ia}(n) = [x_a(n), x_a(n-1), ..., x_a(n-m)]$ are the exogenous inputs and $g(\cdot)$ is a nonlinear function.

In this paper the functions $f(\cdot)$ and $g(\cdot)$ are estimates using a RBF neural network (RBFNN) model.
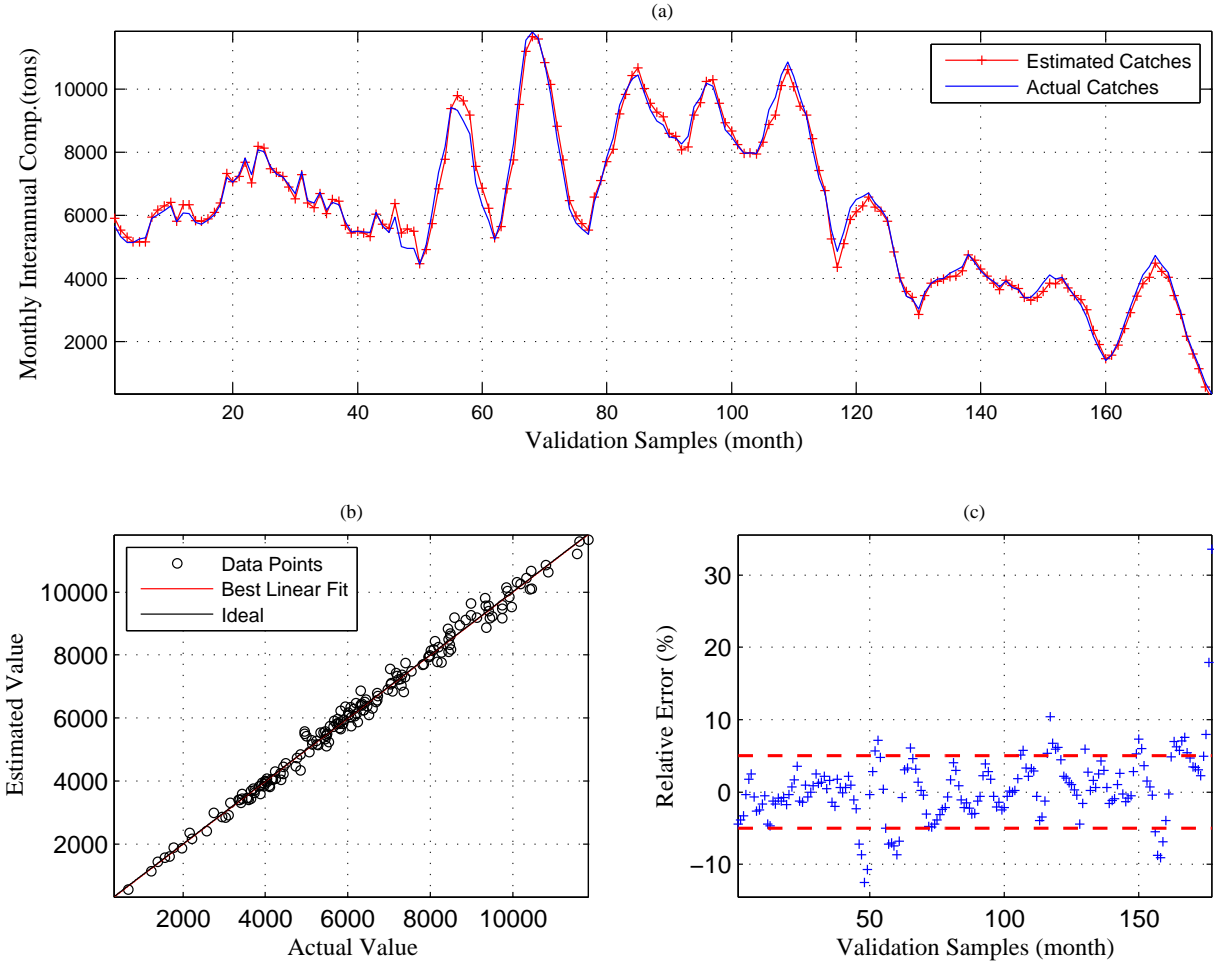
Fig. 3. Three-step-ahead forecasting for inter-annual component

C. RBF neural network model

The output of the RBFNN is obtained as

$$y = \sum_{j=1}^{N_h} b_j \phi_j(\| (z_i - v_{ji}) \|^2) \qquad (8)$$

where N_h is the number of hidden nodes, z notes the regression vector containing $2m$ lagged values, $[b_1, \ldots b_{N_h}]$ represents the linear output parameters, $v = [v_{j1}, v_2, \ldots v_{j2m}]$ denotes the nonlinear parameters, and $\phi_j(\cdot)$ are hidden activation functions, which is given by:

$$\phi_j(\lambda) = \frac{1}{\sqrt{1 + \lambda}} \qquad (9)$$

In order to estimate the linear parameters and the nonlinear parameters of the RBFNN forecaster the separable nonlinear least squares algorithm is used, which is based on least square (LS) method [13] and Levenberg-Marquardt (LM) algorithms [14]. The LS algorithm is used to estimate the parameters b_j, whereas the LM algorithm is used to calibrate the nonlinear parameters v_{ji}. For any given representation

of the nonlinear parameters, the optimal values of the linear parameters are obtained using the LS algorithm as follows:

$$b = \Phi^\dagger X \qquad (10)$$

where X is the desired output patter vector and $\Phi^\dagger$ is the Moore-Penrose generalized inverse of the activation function output matrix Φ [13].

Once linear parameters are obtained, the LM algorithm adapts the nonlinear parameters of the hidden activation functions minimizing mean squared error. Finally, the LM algorithm adapts the nonlinear parameter $\theta = [b_j, v_{ji}]$ according to the following equations [14]:

$$\theta(n + 1) = \theta(n) + \Delta\theta(n) \qquad (11a)$$
$$\Delta\theta(n) = (\xi\xi^T + \mu I)^{-1}\xi^T e \qquad (11b)$$

where ξ represents the Jacobian matrix of the error vector evaluated in θ and e_i is the error vector of the RBFNN for i-patter, I denotes the identity matrix and the parameter μ is increased or decreased at each step of the LM algorithm.
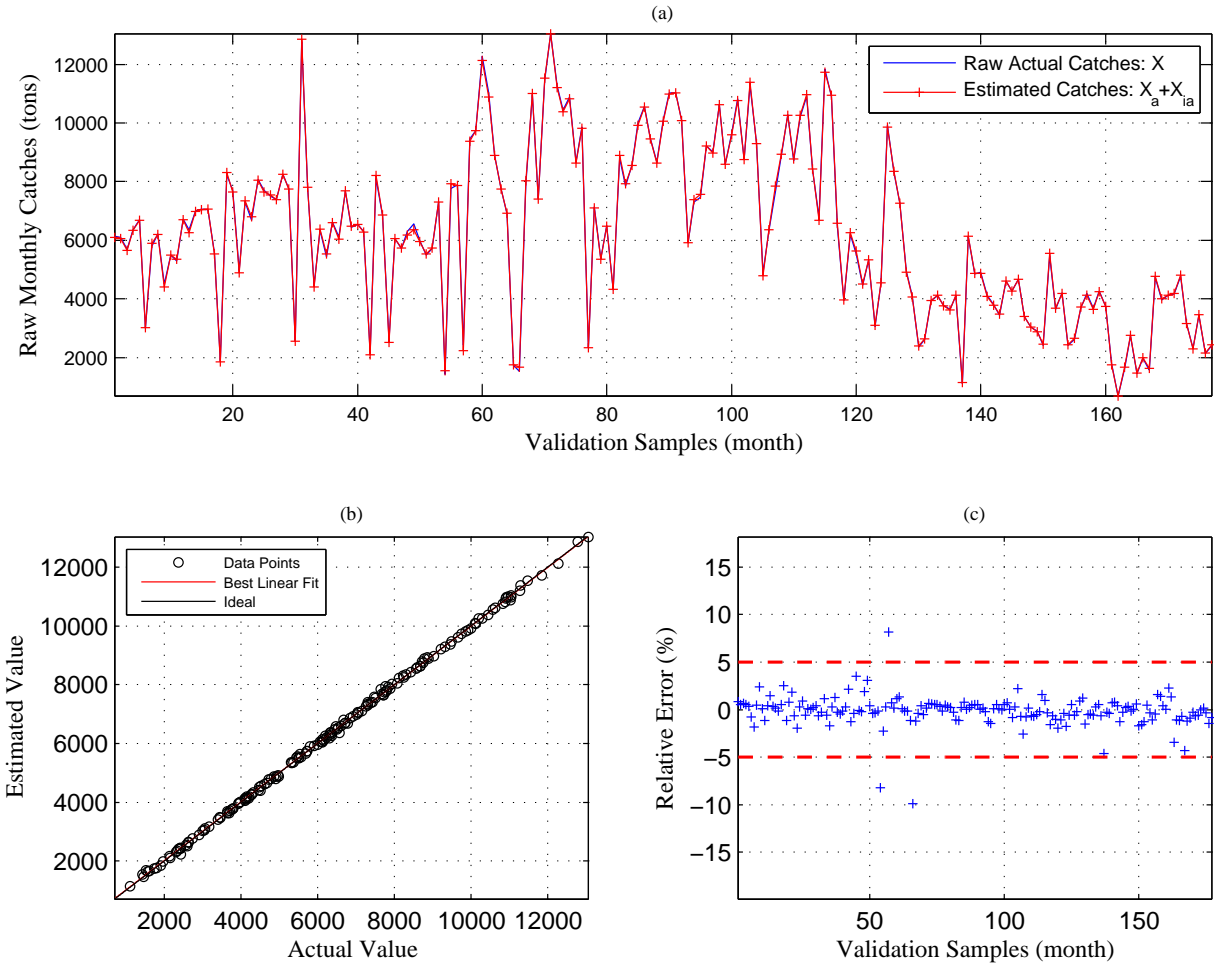
Fig. 4. Three-step-ahead forecasting for raw hake catches vs estimated hake catches

D. Performance Metrics

The forecasting accuracy is evaluated according to the root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and relative error (RE). The forecasting accuracy is better when the values of these measures are smaller. The definition of these metrics are given as follows:

$$RMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(x_i - \hat{x}_i)^2} \quad (12)$$

$$MAE = \frac{1}{M}\sum_{j=1}^{M}|(x_i - \hat{x}_i)| \quad (13)$$

$$MAPE = \frac{1}{M}\sum_{j=1}^{M}|(x_i - \hat{x}_i)/x_i| \quad (14)$$

$$RE = (x_i - \hat{x}_i)/x_i \quad (15)$$

where x_i is the actual value at time i, $\hat{x}_i$ is the forecasted value at time i and M is the number of testing samples.

III. DISCUSSION

Total monthly hake catches off southern Chile were taken from Statistical Fishery Yearbooks (www.sernapesca.cl) for monthly period from January 1963 to December 2008 as shown in Figure 5(a), with a total of $N = 552$ observations. All raw monthly catches data set after subjected to the extraction process of components based on 3-level wavelet decomposition (illustrated in Figure 5), allowed to obtain the inter-annual cycle component and the annual component, which are shown in Figure 5(b) and Figure 5(c); respectively. Once the annual cycle and the inter-annual components were identified, the data set is divided into two parts: a training data set (T=370 observations) and a test data set (M=182 observations).

The training data is firstly used to choose the parameters of the RBFNN models, and the testing data set is used to compute the performance metrics of the models and for validation purposes. The RBFNN was calibrated with $2m = 22$ previous months as input data due to the annual cycle effect of monthly hake catches. Finding the optimal number of hidden nodes is
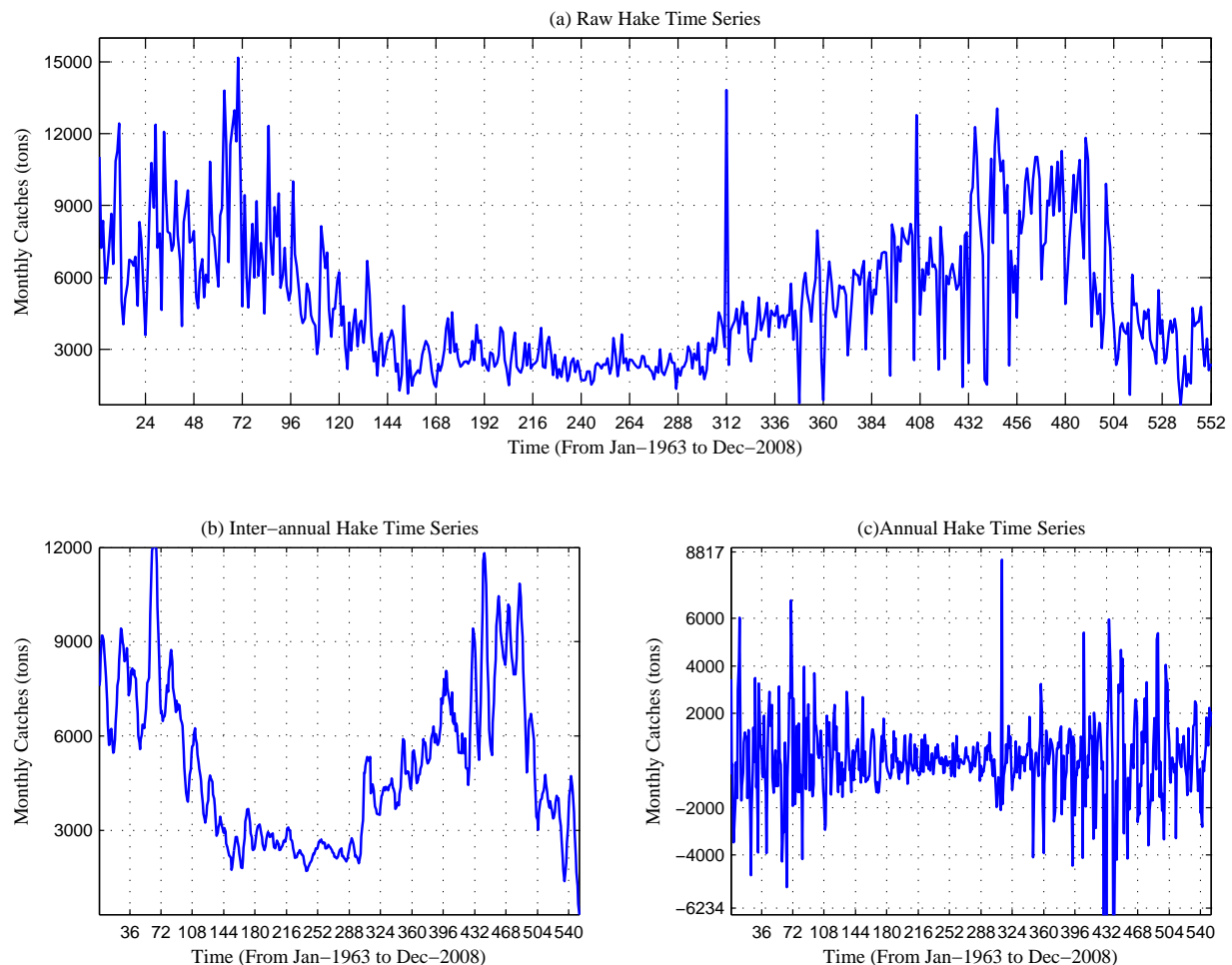
Fig. 5. Reconstruction of hake catches data derived from SWT

a complex problem, but in all our experiments the number of hidden nodes is set to $log(T)$. In the training process overall hidden weights were initialized with the training catches data and the stopping criterion was three iterations. After of the training process, the better architecture was calibrated with 22 input nodes, 6 hidden nodes and one output nodes and is denoted as RBFNN(22,6,1).

Now we present the 3-month-ahead forecasting results obtained with the RBFNN(22,6,1) model during the testing phase, whose results are illustrated in Figures 2, 3 and 4; respectively. Figure 2(a) provides observed annual cycle component versus forecasted annual cycle component, while that Figure 3(a) shows observed inter-annual component versus forecasted inter-annual component. On of other hand, Figures 2(b) and 3(b) show the regression curve between observed components and estimated components. From Figures 2(b) and 3(b) can be seen a good fit of the data to line 1 : 1 with a 98% and 99% of the explained variance for annual component and inter-annual component; respectively.

Figure 4(a) provides raw monthly hake catches data versus forecasted hake catches, whose forecasting behavior is very accurate for testing data with a 99% of explained variance (Figure 4(b)), a MAE of 45 tons and a RMSE of 61 tons, while the explicated variance was of 99%. On the one hand, Figures 2(c), 3(c), and 4(c) depict relative error versus the predicted catches obtained by the RBFNN(22,6,1) model. It can be observed, that an important fraction of the catches tested are acceptable with residuals ranging from $\pm5\%$.

IV. Conclusion

In this paper a 3-step-ahead forecasting strategy for monthly hake catches data set was proposed. The reason of the improvement in forecasting accuracy was due to use multi-scale stationary wavelet decomposition to separate both the annual and inter-annual components of the raw time series, since the behavior of each component is more smoothing than raw data set.

The forecasting strategy was applied to the monthly hake catches of the southern zone of Chile and the results show that 11 previous months of the annual component and 11 previous months of the inter-annual component contain valuable information to explicate a highest variance level.

Finally, the 3-step-ahead RBFNN(22,6,1) forecasting model can be suitable as a very promising methodology to any other marine species of the fisheries industry.

REFERENCES

[1] K. I. Stergiou and E. D. Christou, "Modelling and forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series," *Fisheries Research*, vol. 25, pp. 105–138, 1996.

[2] K. I. Stergiou, "Prediction of the mullidae fishery in the eastern mediterranean 24 months in advance," *Fisheries Research*, vol. 9, issues 1, pp. 67–74, 1990.

[3] J. C. Gutirrez-Estrada, C. Silva, E. Yañez, N. Rodrguez, and I. Pulido-Calvo, "Monthly catch forecasting of anchovy engraulis ringens in the north area of Chile: Non-linear univariate approach," *Fisheries Research*, vol. 86, pp. 188–200, 2007.

[4] J. Gutirrez-Estrada, E. Yañez, I. Pulido, C. Silva, F. Plaza, and C. Borquez, "Pacific sardine (sardinops sagax, jenyns 1842) landings prediction: A neural network ecosystemic approach," *Fisheries Research*, vol. 100, pp. 116–125, 2009.

[5] E. Yañez, F. Plaza, J. Gutierez, N. Rodriguez, M. Barbieri, I. Pulido, and C. Borquez, "Anchovy (engraulis ringens) and sardine (sardinops sagax) abundance forecast of northern Chile: A multivariate ecosystemic neural network approach," *Progress in Oceanography*, vol. 87, pp. 242–250, 2010.

[6] O. Kisi, "Stream flow forecasting using neuro-wavelet technique," *Hydrological Processes*, vol. 22, issues 20, pp. 4142–4152, 2008.

[7] F. K. Nima A., "Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method," *International Journal of Electrical Power & Energy Systems*, vol. 30, issue 9, pp. 533–546, 2008.

[8] Z. Bai-Ling, C. Richard, M. Jabri, D. Dersch, and B. Flower, "Multiresolution forecasting for futures trading using wavelet decompositions," *IEEE Transaction on neural networks*, vol. 12, pp. 765–775, 2001.

[9] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics. Springer-Verlag*, 1995.

[10] G. Nason and B. Silverman, "The stationary wavelet transform and some statistical applications." in *In: Lecture Notes in Statistics: Wavelets and Statistics*, 1995.

[11] J. Pesquet, H. Krim, and H. Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE transactions on signal processing*, vol. 44, pp. 1964–1970, 1996.

[12] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, C. U. Press, Ed. Cambridge, England, 2000.

[13] D. Serre, *Matrices: theory and applications*, Springer, Ed. New York, 2002.

[14] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11(2), pp. 431–441, 1963.