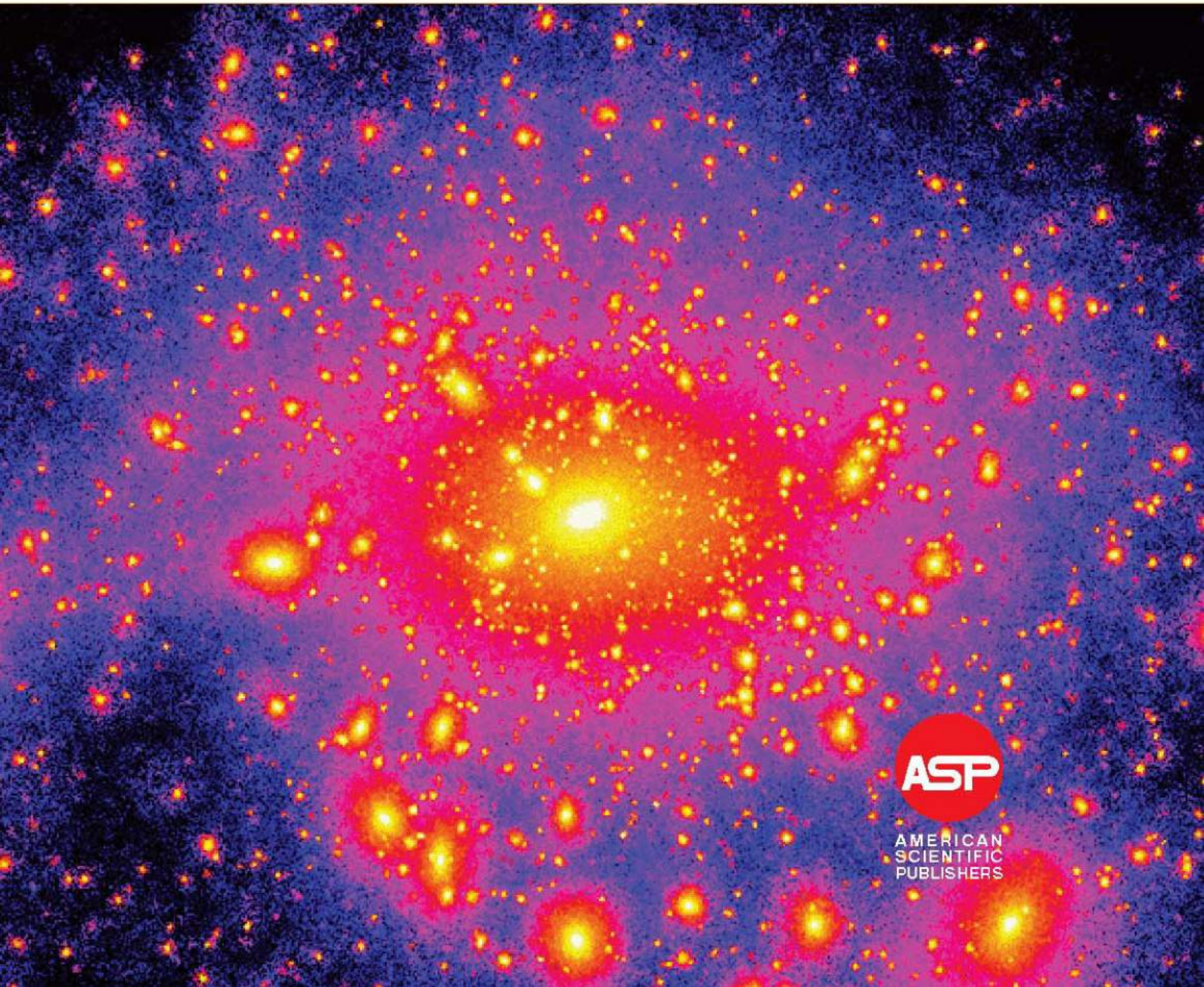
VOLUME 22 • NUMBER 7 JULI 2016

www.aspbs.com/science

Advanced SCIENCE LETTERS

A Journal Dedicated to All Aspects of Scientific Research



ASP
AMERICAN SCIENTIFIC PUBLISHERS

RESEARCH ARTICLE

Advanced Science Letters
Vol. 22, 1860–1864, 2016

Time Series Forecasting Using a Hybrid Principal Component Analysis and Global Ridge-Regression Neural Network Model

Bain Khusnul Khotimah* and Sigit Susanto Putro

University of Trunojoyo Madura, Telang Raya, Madura Island, Indonesia

This research of forecasting the most developed is the time series, which uses a quantitative approach to the data of the past that made reference to forecasting the future. This experiment propose methods to complete time series by reducing the PCA features and improvements Radial Basis Function Neural Network (RBFNN). Where RBFNN is done by combining the functions of RBF and weighting function to produce optimal outcomes. The weights are gotten from regression. Global-ridge regression adds a regulation to give the optimal parameters that produce an optimal weights output. While the feature improvements by using PCA. In the application of research on forecasting the data of SMEs that have a variable with a value of nonlinear show that the method of incorporation RBFNN and global-ridge regression is called (GRNN) have more accurate results compared to some of the methods shown in the value of MAPE, and MSE of all variables data.

Keywords: PCA, Radial Basis Function, Global-Ridge Regression, Neural Network, Time Series.

1. INTRODUCTION

Neural network is an approach used in the forecasting problems. The objective of this study is to investigate the effects in the forecast performance. The ability of the neural network has advantages compared to the statistical models. The advantage is not troubled by the terms and able to predict in the period length of time. Some learning algorithms in neural networks (Artificial Neural Network) developed for time series forecasting. One of them is using the radial basis functions or Radial Basis Function Neural Network (RBFNN), is used for problem resolution time series forecasting and modeling, in addition to the RBF very well when used to solve the problems of non-stationary component and a non-linear. RBFNN work to generate matrix designs where the design matrix requires data input, centroid and activation function. While the activation functions are often used on RBFNN is Gaussian Activation Function, Cauchy and multi kuadratik.[4,6] Design matrix generated by RBFNN is used to found parameter. Parameter estimation method used is the global ridge-regression method. As for the dimensions of the data used Principal Component Analysis (PCA) is a statistical method used in order to reduce the dimensions of the input with the minimum loss of information. Some researchers have combined PCA with neural network. PCA can also be combined with neural network as input variables RBFNN to reduce the problem of forecasting.[3,4]

This research base on the model-RBPNN PCA, PCA is used to reduce the dimension input into some of the major components of a smaller dimension. PCA algorithm designed to approximate a pattern of data on high-dimensional space with sub low dimensional space which spans the principal Eigen Vector of covariance matrix data. In this way the data distribution can be expressed and reconstructed with the principal eigenvectors and eigenvalues vectors. The next major component of the PCA results are used as input in to RBPNN. Principal Component Analysis (PCA) is a statistical method used in order to reduce the dimensions of the input with the minimum loss of information. Some researchers have combined PCA with neural network to reduce the input variables in forecasting problems. Parameters of input dimensions are used to determine the best weight for modeling Radial Basis Function Neural Network (RBFNN). RBFNN has the advantage of the algorithm is simpler and faster computing when compared with algorithms Backpropagation, recurrent and so forth. RBFNN can use to establish a relationship noisy, interpolation, classification and estimation functions.[3,11]

This study uses the modeling radial basis function has been performed for time series data with activation function Gaussian with the method of global ridge-regression neural network (GRNN) with PCA combination, and while forecasting RBF using forecasting methods of constructive learning and to estimate the weight by the method global ridge-regression to the value of the variant using three definitions, namely standard

* bain@trunojoyo.ac.id

1936-6612/2016/22/1860/005 doi:10.1166/asl.2016.7050

deviation, the average distance from the center of the cluster and the maximum distance from the cluster center forecasting GRNN.

2. TIME SERIES PREDICTION AND NEURAL NETWORKS

Forecasting time series is used to analyze the data a variable observations were prepared following the time sequence. It is very necessary, especially if we only have a data set of historical data and do not understand the factors that affect the system. Various methods of forecasting has been widely used, and produce a diverse accuracy. Attempts to get the most accurate in forecasting continued. One of the efforts is by leveraging the capabilities of the architecture and methods in the neural network. Elman recurrent neural network in the neural network used in situations when there is a connectedness data depends on the time. Recurrent neural network is a modification of the feed forward neural network which is very popular, and appropriately used for forecasting time series.

As an illustration, FFNN architecture with one hidden layer (hidden units) can be seen in Figure 1. Suppose the process of m-variant $\{\mathbf{Z}t\}$ consists of and n observations. Model FFNN for multivariate time series data of the response is as follows.

$$Y = b^o + \sum_{j=1}^{q} w_j^o f_j^h (X w_j^h + b_j^h) + u \tag{1}$$

With Y_i, Matrix input X_i, vector of parameter w_i, and vector error u. Estimation of the model parameters is done with method backpropagation. If Y is output data and $f(X; w)$ is the output of the model NN, then the parameter vector w estimated by minimizing the function

$$E(w) = (Y - f(X; w))'(Y - f(X; w)) \tag{2}$$

as is done in a non-linear regression. To minimize the function *E(w)*, backpropagation method using a linear approximation of a function of the error is

$$E(w + \Delta w) \approx E(w) + \Delta w^l E'(w) \tag{3}$$

Weights updated through

$$\Delta w = -\eta E'(w), \quad \eta > 0 \tag{4}$$

with η is a coefficient learning (learning rate).

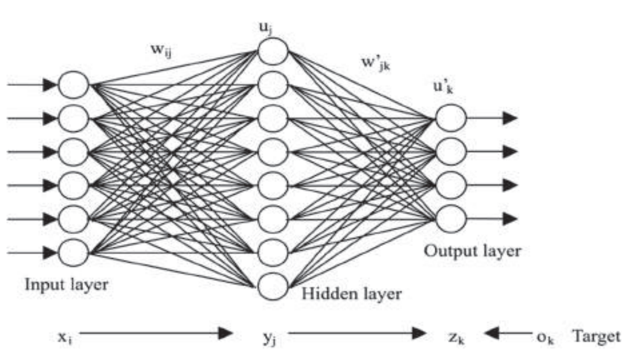


Fig. 1. Architecture RBPNN.

3. PCA-RBPNN

Principal Component Analysis (PCA) is an analysis to reduce dimensional variables in multivariable of data, data compression, pattern recognition and statistical analysis. PCA process is used to approximate original high-dimensional data space with lower dimension in the sub space spanned by a principal eigenvector of the covariance matrix data. In this way, the distribution of the data can be represented and reconstructed with the Principal Eigen Vectors and Eigen Values related. PCA goal is to reduce the dimensions of the variable input data into the main component of a smaller dimension with minimum information loss, in which the main components are not correlated with each other.

Projections on the PCA is a representation of the data set X in the form of orthonormal eigenvectors of the data covariance matrix X. Covariance matrix is the correlation between the variables in the data set X. PCA is the process of getting orthonormal eigenvectors of the covariance matrix as a base to be transformed into the data space the new one. Eigenvectors can be regarded as the original basis for the multi-dimensional data X. Largest Eigen Value of the covariance matrix is the smallest correlation between variables in the data space. Following PCA will seek the projected variables that are not correlated.

The main component is a linear combination of $Y_1, Y_2, \ldots, Y_p$ uncorrelated with the largest variant. The first principal component is a linear combination of a1'X with variants Var $(a_i'X)$, the largest in $a1'a1 = 1$ and the second main component is $a1'X$ linear combinations of the variants Var $(a2'X)$, the largest in $a2'a2 = 1$ and Cov $(a_1'X, a_2'X) = 0$. Since the i-th principal component is a linear combination of the variants $ai'X$ Var $(a_i'X)$, the largest in $ai'ai = 1$ and Covariant $(a_i'X, a_k'X) = 0$, $(k < i)$. Projections on the PCA is a representation of the data set X in the form orthonormal Eigen Vector of covariance matrix of data X. Covariance matrix is the correlation between the variables in the data set X. PCA is the process of getting orthonormal eigenvectors of the covariance matrix as a base to be transformed into the data space the new one. With the principal components analysis we would reduce the observational data into multiple sets of data so that information from all the data we can absorb optimally. Thus the principal component analysis can be viewed as the transformation of the $X_1, X_2, \ldots, X_p$. For example $X_1, X_1, \ldots, X_p$ has a variance-covariance matrix $\sum = (\sigma_{ij}^2)$, $i = 1, 2 \ldots, p$: $j = 1, 2, \ldots, p$ and the Σ has eigen values $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$.

The first Principal Component expressed by PC_1 contains the greatest amount of total variation of the data. PC_1 as linear combinations of the variables Xi and PC_1.

$$\begin{aligned} &X_i; \quad i = 1, 2, \ldots, p \\ &PC_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \end{aligned} \tag{5}$$

where a_{1i} selected, so as to maximize the ratio of variance PC1 to the total variance, with a barrier that $\sum a_{1i}^2 = 1$.

The principal component regression formation through principal component analysis, there are two ways. First, the establishment of the main components based on the covariance matrix. Secondly, the formation of the main components is based on the correlation matrix. Through the data source to be searched X_{nxp} variance covariance matrix Σ where elements are

$$S_{jk} = \frac{1}{n-1} \sum_{j=1}^{p} (X_{ij} - X_j)(X_{ik} - X_k) \tag{6}$$

where covariance matrix is searched with Eigen Value λ_i $i = 1, 2, \ldots, p$, optimate determinan are calculated: $|S - \lambda_i I| = 0$.

Eigen Vectors calculated by an equation:

$$Se_i = \lambda_i e_i \quad \text{where } i = 1, 2, \ldots, p \tag{7}$$

Projections on the PCA is a representation of the data set X in the form orthonormal eigen vector of covariance matrix of data X. covariance matrix is the correlation between the variables in the data set X. PCA is the process of getting orthonormal eigenvectors of the covariance matrix as a base to be transformed into the data space the new one. Each group of data has a major component, but the PCA will work well if the data group Gaussian distribution. The PCA always use the form deviations of the mean for any data has been reduced by the mean or often referred to zero-mean. Zero mean probability distribution that fully describe variants is a Gaussian distribution.[10]

3.1. The Hybrid PCA, SOM AND GRNN Methodology

Algorithm Self Organizing Map (SOM) is an artificial neural network method introduced by Professor Teuvo Kohonen in the 1980s. The combination SOM with the model PCA-RBPNN is clustering at the PCA to determine the most dominant major components for each variable data. The main component produced by PCA will then be input SOM to determine the cluster's center and RBPNN input. Suppose $X \in Rm \times n$ known to the data, the PCA process will produce major components $Y \in Rr \times m$ generated. SOM algorithm is used to partition Y become r cluster.

Before the training of human resources, the element weight vector is initialized random from center input. Each input $y(t)$ is initialized as weights randomly chosen as the center with the initialization sequence generated random permutation. Thus every input elected as the center will be the Best Matching Unit (BMU) early, then used to search for BMU on the next iteration. In the Gaussian distance function using the input position and center the randomnya sequence has been determined in the initialization phase center. In SOM training, the first time calculated the Euclidean distance between *dj* the input vector *y(t)* and all the weight vectors center *mi(t)* and then determined the winner unit c with the closest weight vector of the input vector. Unit c winner is determined by the following equation:[8, 12]

$$c = \arg\min\{d_j\} \quad d_j = \|y(t) - m_i(t)\| \tag{8}$$

Vector weighs updated by the equation with rc position, center position to the unit-it) learning rate, $\sigma(t)$ the width of the Gaussian function.

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t, \|r_c - r_i\|)[y(t) - m_i(t)] \tag{9}$$

SOM is a preprocessing process for the determination of the initialization RBFNN center and the target of RBPNN. Suppose that the initial data input is known $X \in Rm \times n$, where m is the dimension or variable and n is the number of data samples. Since, The PCA process reduced the size of the input data and produce variable $Y \in Rr \times m$, where r is the number of major components. The main component of the PCA results clustering and produces measuring c center and cluster size $r \times 1$. Then C and T used as initialization for RBPNN. Input data is a major component PCA results with cumulative sufficient to explain the overall diversity data of size $r \times m$.

Mathematically, RBPNN with input vector x will produce the actual value for i-th output neuron y_i^a expressed as the following equation:

$$\begin{aligned} y_i^{\alpha} &= \sum_{k=1}^{M} w_{ik} h_k(x) \\ h_k(x) &= \sum_{i=1}^{nk} \phi_i(x, c_{ki}) \\ &= \sum_{i=1}^{nk} \phi_i(\|x - c_{ki}\|), \quad k = 1, 2, \ldots, M \end{aligned} \tag{10}$$

As for $\phi_i(\cdot)$ is a kernel function that is generally Gaussian kernel function

$$\phi_i(\|x - c_{ki}\|_2) = \exp\left(-\frac{\|x - c_{ki}\|_2^2}{\sigma_i^2}\right) \tag{11}$$

where σ_i is the Gaussian kernel function parameters. The fastest and simplest method for determination RBPNN spread by using the election σi given indimana σ_i adalah parameter Kernel Gaussian Function.

$$\sigma = \frac{d_{\max}}{\sqrt{K}}$$

with $d_{\max}$ is the maximum Euclidean distance of a set of training and K is the number a total of a set of training is the maximum Euclidean Distance of a set of training and K is the number a total of a set of training.[7]

3.2. Weighted on Ridge Regression Modeling GRNN for Time Series

Network is statistical modeling to the data series. One form of neural network model of radial basis function network (RBF) called RBFNN. RBF model is comprised of input layer, hidden layer and output layer. While the input layer and hidden layer has a number of units that varied in many ways determine the number of units in the hidden layer, variant, weight, parameter estimates on neural network models and suitability criteria RBF models on modeling of time series data. In this work, we use a set of methodologies that can be used assessing statistical modeling for forecasting the RBF models and Using the global ridge-regression weighting algorithm output as a suitable candidate on the model of RBF. Estimating the exact weight will produce output function with minimal error and will add a regularization parameter that is minimize SSE (sum of square error) to obtain the cost function Weight-decay together with the ridge regression.

$$C = \sum_{i=1}^{p} (\hat{y} - f(x_i))^2 + \lambda \sum_{j=1}^{m} w_j^2 \tag{12}$$

Global ridge-regression has one parameter for all base function, so we get the equation

$$\sum_{i=1}^{p} f(X_i) h_j(X_i) + \lambda w_j = \sum_{i=1}^{p} \hat{y} h_j(X_i) \tag{13}$$

Where m to the equation $1 \le j \le m$, that will generate weight.

In this study SOM clustering method is used as the center of the selection procedure. So, the selection of the number of

basis function (or unit hidden) RBF network used constructive learning. Global ridge-regression method is used to determine the weight of the coating. The relevance between the number of hidden units with the output from the criteria calculated MAPE, MAD and MSE. Rbf Design program used to generate the design matrix for RBFNN with the input, center, variance, type bias unit functions and options that remain. To get closer to the output data is actually used program ridge regression that can improve matching. Ridge regression or weight-decay has two forms of global ridge with a single parameter and local ridge with the parameter m to m base function. Optimal matching of data by preventing over-fitting by adding a positive regularization parameter λ on the SSE so that the resulting cost function as follows:

$$c = \sum_{i=1}^{p}(f(x_i) - y_i)^2 + \lambda \sum_{j=1}^{m} w_j^2 \tag{14}$$

Global ridge-regression produce optimal weight vector with the normal equation as follows:

$$\hat{w} = (H^T H + \lambda I_m)^{-1} H^T y \tag{15}$$

H is the matrix design, Y is the output vector with p rows and n columns (n is the dimension of the output space), l is the alleged value of regularization parameter positive λ, options is the choice of the method used to forecast error on the test data and U are matrices penalty. Optimization is done using the initial allegation and one preferred method to minimize error in the validation data set in the future. Methods to minimize the error used to obtain optimal value of λ.

4. EXPERIMENTAL RESULT

RBPNN is the initial process for determining the center and clustering for each input y_i. PCA-GRNN the model, the model used for the determination of SOM clustering center and clustering initialization input of RBPNN. In the process of SOM randomly selected sequence initial center of the input vector y_{ij} as initialization center. With a number 9 permutations of random sequence initial center by random data sequence. Number of clusters is determined by the amount of the main components resulting in PCA. SOM end of the process will produce center (c_{ij}, i, $j = 1, 2$) and clusters (Ti, $i = 1, \ldots, N$) of the overall input data RBPNN. Center CIJ and further Ti cluster used as initialization RBPNN center and the target for the prediction X competitive layer. SOM average error is ≤ 0.1, with parameters $\alpha(t)$ with values $0 \leq \alpha(t) \leq 0.5$. SOM clustering results with used as a center for RBPNN initialization.

Data for the test model is a time series data availability of oil in POM Gasoline Camplong in Pamekasan. Recording a report on reservations, the number of sales, price, rate of price increases, inventories, the number of buyers every day and so is a component of the system is the most important information and basis used by company managers. Due to this marketing managers can find an important issue that is by analyzing it to identify the pattern of fuel sales data. 7869 experimental data consists of data points daily average of 2010, 2011, 2012 and 2013. Based on the measurement data of 2010–2013 there were nine parameter measurements are contained in the original data measurement $x_1, x_2, \ldots, x_9$. From the above explanation can be concluded that despite the variable input has been reduced by the PCA which led to the loss of some initial data information, but the predictive ability for x_1 and x_2 with PCA-GRNN still accurate. From the experimental results for GRNN input dimensions 9×9 whereas the PCA is able to be reduced to 9×2. This will certainly also resulted in the clustering process by SOM will be faster since the size of the center will be a smaller size of 2×2. This also resulted in the first hidden layer structure on GRNN will be simplified. Running time comparison results generated PCA with

Table I. Training results in a comparison test.

	SME			MAPE		
	RBPNN	PCA-RBPNN	PCA-GRNN	RBPNN	PCA-RBPNN	PCA-GRNN
100	0.1452	0.0654	0.0354	0.0854	0.0754	0.0554
200	0.1742	0.2371	0.1371	0.2371	0.1371	0.1271
300	0.1908	0.2495	0.1249	0.3495	0.2495	0.1495
400	0.2020	0.2711	0.1212	0.3711	0.2711	0.1711
500	0.2357	0.4310	0.2310	0.4310	0.3310	0.2310
600	0.7939	0.6003	0.4003	0.6023	0.6003	0.5003
700	0.8163	0.7406	0.6406	0.9406	0.8406	0.7406
800	0.5388	0.4775	0.5775	0.9775	0.8775	0.7685
Mean	0.3871125	0.384063	0.2835	0.499313	0.422813	0.342938

Table II. Testing process results in a comparison test.

	SME			MAPE		
Record data	RBPNN	PCA-RBPNN	PCA-GRNN	RBPNN	PCA-RBPNN	PCA-GRNN
100	0.1251	0.0854	0.0654	0.0754	0.0654	0.0454
150	0.1541	0.1371	0.1271	0.2171	0.1261	0.1162
250	0.1818	0.2295	0.1349	0.3495	0.2495	0.1394
350	0.2010	0.2561	0.1413	0.3711	0.2711	0.1632
450	0.2251	0.2315	0.2118	0.4310	0.3310	0.2210
550	0.7639	0.5002	0.4003	0.6023	0.6003	0.5403
600	0.7163	0.6404	0.5405	0.9406	0.8406	0.7241
700	0.8388	0.5785	0.5675	0.9775	0.8775	0.7585
Mean	0.4007	0.3323	0.2736	0.4955	0.4201	0.3385

some other methods. At the trial stating the time required for data extraction train, tUji stated time for extraction of test data during the process of forecasting data measured at the time of computation.

According to Tables I, II would be seen for a change of epoch 100 will get the smallest MSE value is MSE $\leq$ 0, 1. It can be concluded that the greater constraints epoch which included the value of the error will be smaller. But this does not apply to the epoch value $\geq$100 because when the epoch of 100 have got the smallest error value, but if the epoch which included more than 100 then the system will continue to process until it reaches a maximum epoch but the resulting error will increase in size will simplify *Global Ridge-Regression* Neural Network.

5. CONCLUSIONS

The linear combination GRNN models with the PCA and the nonlinear ANN models are used jointly, aiming to capture different forms of relationship in the time series data. The hybrid models takes advantage SOM for clustering, PCA and GRNN of the unique strength of linear and nonlinear modeling. By testing the scenario differences due to differences in the amount of data teller training and testing it with more data than the greater the resulting error. The different number of hidden neurons in each variable initialization is also influenced by the weight of the SOM competitive, where it is the number of neurons that lead to differences in input current and the time of testing. Thus the weight center in the second hidden layer RBPNN to ROLS will be much more modest if compared with RBPNN, so that the process of projection data or predictions will be much faster. Weights end RBPNN center will be used for prediction in the sample and the sample in the process of testing out RBPNN GRNN combination with PCA model is the best model currently used for test data by calculating the size of the error using the MSE.

References and Notes

1. A. Jain and A. M. Dan Kumar, *Applied Soft Computing* 2, 585 **(2007)**.
2. M. Khashei and M. Dan Bijari, *Expert Systems with Applications* 7, 479 **(2010)**.
3. Z. X. Li and H. G. Dan Guang, *System Engineering* 1, 167 **(2007)**.
4. W. Z. Lu, W. J. Wang, X. K. Wang, S. Y. Yan, and J. C. Dan Lam, *Environmental Research* 96, 79 **(2004)**.
5. D. Peng, Z. Yi, J. C. Lv, and Y. Dan Xiang, *Neurocomputing* 71, 1748 **(2008)**.
6. V. M. Rivas, J. J. Marelo, P. A. Castillo, M. G. Arenas, and J. G. Dan Castellano, *Information Science* 165, 207 **(2004)**.
7. D. S. Huang and W. B. Dan Zao, *Applied Mathematics and Computation* 162, 461 **(2005)**.
8. D. S. Huang and J. X. Dan Du, *IEEE Transactions on Neural Network* 19, 2009 **(2008)**.
9. Z. X. Li and H. G. Dan Guang, *System Engineering* 27, 167 **(2007)**.
10. J. Vesanto and E. Dan Alhoniemi, *IEEE Transaction on Neural Networks* 11, 586 **(2000)**.

Received: 10 December 2015. Accepted: 12 December 2015.