CHAPTER III

APPLICATIONS OF INCOMPLETE TIME-SERIES PREDICTION

In this chapter, we firstly describe the concept of incomplete time-series prediction. Then, we describe the construction of both FI-GEM and RMD-FSE networks for improving the prediction accuracy of incomplete time-series prediction.

In this work, we focus on periodic time-series data. The periodical time-series data may be seasonal or cyclical. The data can be modelled as follows: Let x_t be the value of a system at time t. The value of x_t is given by $x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-k})$, where k previous samples are used in the modelling. We denoted k to be the cycle of the time-series data. The training data are partitioned and arranged in a time-series form with a window size of k. Each set of data is, then, stacked to form an input matrix A as shown in equation (3.1).

$$A^{Training} = \begin{bmatrix} x_1 & x_2 & \cdots & x_k \\ x_2 & x_3 & \cdots & x_{k+1} \\ x_3 & x_4 & \cdots & x_{k+2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N-k+1} & x_{N-k+2} & \cdots & x_N \end{bmatrix} \tag{3.1}$$

Suppose x_t is a missing datum. Obviously, the number of x_t can appear diagonally from 1 to k times. Let the number of appearances of x_t in A be k'.

For example, there is a periodical time-series data, whose cycle is four.

$$Dat = \begin{bmatrix} 0.5 & 0.3 & 1 & 0.6 & 0.7 & 0.2 & 1.1 & 0.5 & 0.4 & 0.4 & 1.2 & 0.4 \end{bmatrix} \tag{3.2}$$

We suppose that the data is missing at time ith. The index ith are fourth and sixth. Symbol "?" denote the missing data. The inputs are partitioned to the training set as follows:

$$A^{Input} = \begin{bmatrix} 0.5 & 0.3 & 1 & ? \\ 0.3 & 1 & ? & 0.7 \\ 1 & ? & 0.7 & ? \\ ? & 0.7 & ? & 1.1 \\ 0.7 & ? & 1.1 & 0.5 \\ ? & 1.1 & 0.5 & 0.4 \\ 1.1 & 0.5 & 0.4 & 0.4 \\ 0.5 & 0.4 & 0.4 & 1.2 \end{bmatrix} \tag{3.3}$$

$$O^{Output} = \begin{bmatrix} 0.7 \\ ? \\ 1.1 \\ 0.5 \\ 0.4 \\ 0.4 \\ 0.4 \\ 1.2 \end{bmatrix} \tag{3.4}$$

A training set and a target set are incomplete in equation (3.3) and (3.4). In general, when we trained a neural network, the input and target should be completed. Thus, filling in the estimating missing value is needed, we should not ignore those missing.

Each $\hat{x}_t$ is modeled as $x_t + \epsilon$. When we considered $x_t + \epsilon$, which is like the noisy data problem. The problem is to determine the technique that produces $\hat{x}_t$ such that $(\hat{x}_t - x_t)^2$ is minimized. Hence, we should consider a fill-in technique reducing the noise effect of each x_t.

We have proposed the applications of time-series prediction as follows

3.1 FI-GEM Network

We consider several standard techniques for filling in the missing data in our experiments. We used cubic smoothing spline interpolation and the imputation based on EM algorithm [3] in our preliminary solution. These techniques are applied to the same data set with some missing data. We have to train the individual networks by supervised neural networks multilayer perceptron(MLP) with extended Kalman filtering [25] by Singhal. The ensemble construction [24] Perrone is used for the combination of the individual networks. We name this type of network *Fill-In - Generalized Ensemble Method (FI-GEM)*networks. The structure of FI-GEM network is shown in Figure 3.1.

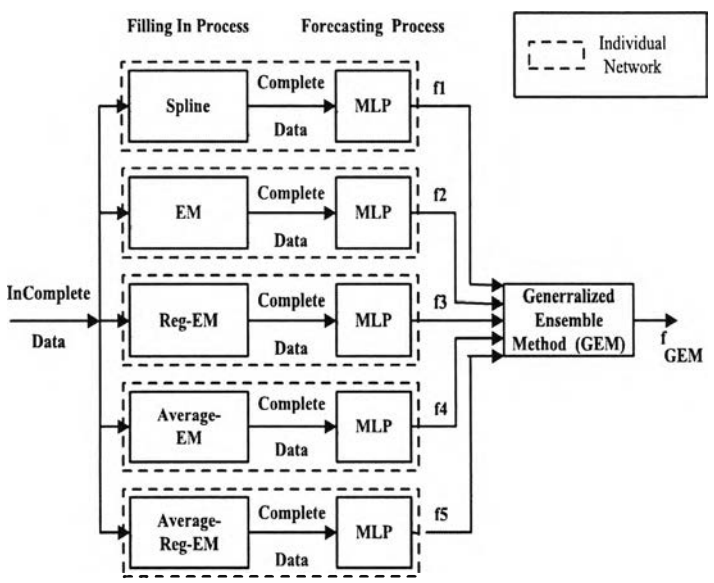


Figure 3.1: FI-GEM Network Structure

A. Training Neural Network : Let $o_j^{(q)}$ be the output signal of j^{th} neuron in the q^{th} layer induced by the presentation of an input pattern t, $\mathbf{x}_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,N})$, and $w_{i,j}^{(q)}$ the connection weight coming from the i^{th} neuron in the $(q-1)$ layer to the j^{th} neuron in the q^{th} layer. Assume that $\mathbf{x}$ is an augmented vector, i.e. $x_0 = 1$. Then $o_j^0 = x_j$, $o_j^{(q)} = g(net_j^{(q)})$, and

$$net_j^{(q)} = \sum_{i=0}^{n_{q-1}} w_{ij}^{(q)} o_i^{(q-1)} \tag{3.5}$$

when n_{q-1} is the number of nodes in the $(q-1)$ layer. The activation functions used in each neuron of the hidden layer and of the output layer are a sigmoidal function $g(a) = \frac{1}{1+exp(-a)}$ and an identity function $g(a) = a$, respectively.

B. Learning Algorithm : Suppose there are M neurons in the output layer, T input patterns, and Q layers. Let $\mathbf{w}_j^{(q)} = [w_{1,j}^{(q)}\ w_{2,j}^{(q)}\ \ldots\ w_{n_{q-1},j}^{(q)}]^T$ be the weight vector of neuron j in the q^{th} layer, W a set of all $\mathbf{w}_j^q$, and $\mathbf{y}_t = [y_{1,t}\ y_{2,t}\ \ldots\ y_{M,t}]^T$ the target of input pattern $\mathbf{x}_t$. For a given training set $\{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \ldots, (\mathbf{x}_T, \mathbf{y}_T)\}$, In the preliminary solution of this work, we use the extended Kalman filtering (EKF) algorithm [25] to train the network by minimizing the sum-squared error.

$$E(\mathbf{W}) = \frac{1}{2} \sum_{t=1}^{T} \sum_{j=1}^{M} (o_j^{(Q)} - y_{j,t})^2 \tag{3.6}$$

The training process will stop if the mean square error (MSE) is less than or equal to the specific tolerance.

$$MSE = \frac{1}{T} E(\mathbf{W}) \tag{3.7}$$

C. Neural Network Prediction : A combination of many neural networks of the same type significantly shows the improvement of the prediction performance. Ensemble

networks consist of independently trained neural networks which are combined as a single master network. The network is used as the second estimating step. The input to each sub-network is the output from each missing data fill-in technique.

3.2 RMD-FSE Network

For this network, we still present an approach that uses several EM-based algorithms and a smoothing spline interpolation to fill in the missing data values like FI-GEM network. Each individual network, which is unlike FI-GEM network, is one that uses a Finite Impulse Response (FIR) model [26] to perform the prediction. We denote this approach as a *reconstructed missing data-finite impulse response selective ensemble (RMD-FSE)* network. The RMD-FSE network has two parts, the individual FIR networks and the master network used to integrate the outputs.

A. Individual Networks : Each individual network is a layer-by-layer fully connected feed-forward network modeled with tapped-delayed synapses. Suppose the linking synapse between node i of layer l and node j of layer $l+1$ with a delay of T_l is

$$[w_{ji}^l(0)\ w_{ji}^l(1)\ \ldots\ w_{ji}^l(T_l)]^T,$$

and the input at time n with T_l-delay is $[x_i^l(n)\ x_i^l(n-1)\ \ldots\ x_i^l(n-T_l)]^T$. The FIR filter forms a weighted sum of past values of its input. The neuron j of layer $l+1$ receives the filtered inputs $s_j^{l+1}(n)$ as

$$\begin{aligned} s_j^{l+1}(n) &= \sum_i \{w_{ji}^l(0)x_i^l(n) + w_{ji}^l(1)x_i^l(n-1) \\ &\quad + \ldots + w_{ji}^l(T_l)x_i^l(n-T_l)\}, \end{aligned} \tag{3.8}$$

and then passes $s_j^{l+1}(n)$ through a sigmoidal nonlinear function f as

$$x_j^{l+1}(n) = f(s_j^{l+1}(n)). \tag{3.9}$$

Neurons are arranged in layers to form a network in which all connections are made with synaptic filters. During the training of the network, the value of each weight is adjusted in order to minimize the specified squared error $e(n)$ by using the temporal backpropagation-type algorithm [26].

B. Master Network : The outputs of all individual neural networks are combined by the genetic algorithm-based selective neural network ensemble method (GASEN) [27]:

$$x_{RMD-FSE}(n) = \sum_{i=1}^{N_{RMD-FSE}} \alpha_i x^{(i)}(n),$$

where $N_{RMD-FSE}$ is the number of individual networks, $x^{(i)}(n)$ is the output value of the ith network, and α_i is the weighting parameter for the output from the ith network. The values of α_i must satisfy the constraints of $0 \leq \alpha_i \leq 1$ and $\sum_i \alpha_i = 1$. Choosing the actual values of α_i is an optimization problem because it cannot be known *a priori* which network will produce a more reliable estimate. Let α be the vector such that α_i is its ith element. The parameter vector α can be found by minimizing

$$E_\alpha = \alpha^T \mathbf{C} \alpha$$

where $\mathbf{C}$ is the correlation matrix of the errors from the network predictors $x^{(i)}$ and $x^{(j)}$.

A genetic algorithm is used to search for a solution of E_α by defining $h(\alpha) = \frac{1}{E_\alpha}$ as the fitness function which is to be maximized. The components of a candidate solution α may violate the constraints during its evolution. Therefore its elements α_i should be normalized to $\alpha_i / \sum_j \alpha_j$ at each generation. The networks which are associated with α_i less than a threshold value λ, which is zero, will be excluded.

3.3 Extended RMD-FSE Network

We extended our work [21] to extended RMD-FSE network [22]. Six fill-in methods, viz. cubic smoothing spline interpolation, k-segment principal curves, Expectation maximization (EM), regularized EM, average EM, and average regularized EM, are simultaneously employed in a first step for reconstructing the missing values of time-series data. A set of complete data from each individual fill-in method is used to train a FIR neural network to predict the time series like [21]. The outputs from individual networks are combined by a selective ensemble method in the second step. Experimental results show that the prediction made by the proposed method is more accurate than those predicted by neural networks without a fill-in process or by a single fill-in process. The improvement of performance of master network is explained as follows:

A. Master Network Suppose we use an ensemble of N networks to predict a sample value $f_{RMD-FSE}(n)$. The overall ensemble output can be any function of the individual network outputs. The simplest such function is to combine the individual outputs as a weighted sum:

$$x_{RMD-FSE}(n) = \sum_{i=1}^{N} \alpha_i x^{(i)}(n), \tag{3.10}$$

where $x^{(i)}(n)$ is the output value of the ith network, with an associated weighting parameter α_i. Intuitively, the weights should satisfy $\sum_{i=1} \alpha_i = 1$ because each network produces a prediction of the same sample $x(n)$. Additionally, since the predicted values should have the same sign as the sample, it is reasonable to expect $0 \leq \alpha_i \leq 1$, for all i.

The weight should be set so that if the jth network is more reliable, its associated weight α_i should be larger. Choosing the actual values of α_i is an optimization problem based on the observed mean square error of individual network. Let $\mathbf{C}$ is the correlation matrix of the errors from the network predictors $x^{(i)}$ and $x^{(j)}$; the (i, j)th element of $\mathbf{C}$

is given by

$$c_{ij} = \frac{1}{|\Gamma|} \sum_{k \in \Gamma} e^{(i)}(k) e^{(j)}(k) \tag{3.11}$$

where Γ is the training set and $e^{(j)}(k)$ is the error of Network j produced in response to the kth input of the training set. The diagonal terms of $\mathbf{C}$ are the mean square errors of the individual networks while each off-diagonal term is a pairwise correlation of the corresponding networks. In the following we write c_{ii} as c_i^2. Since the ensemble output should have a better performance than the individual networks, a criteria for finding α is to minimize the ensemble output mean square error:

$$E_\alpha = \alpha^{\mathbf{T}} \mathbf{C} \alpha. \tag{3.12}$$

Some insight about the choice of the weights and the output mean square error can be gained by considering the simple case when $N = 2$. The solution for α is to be found by minimizing

$$E_2 = \begin{bmatrix} \alpha_1 & \alpha_1 \end{bmatrix} \begin{bmatrix} c_1^2 & c_{12} \\ c_{12} & c_2^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}. \tag{3.13}$$

Since $\alpha_2 = 1 - \alpha_1$, we can differentiate E_2 with respect to α_1 to solve for α_1 as :

$$\alpha_1 = \frac{c_2^2 - c_{12}}{c_1^2 + c_2^2 - 2c_{12}}. \tag{3.14}$$

It follows that

$$\alpha_2 = \frac{c_1^2 - c_{12}}{c_1^2 + c_2^2 - 2c_{12}}. \tag{3.15}$$

With this set of weights, the ensemble output has a mean square error of

$$\begin{aligned} E_2 &= \alpha_1^2 c_1^2 + \alpha_2^2 c_2^2 + 2\alpha_1 \alpha_2 c_{12} \\ &= \frac{c_1^2 c_2^2 - (c_{12})^2}{c_1^2 + c_2^2 - 2c_{12}} \end{aligned} \tag{3.16}$$

This set of weight illustrates that when Network 1 is realiable, i.e., $c_1^2 \leq c_2^2$, its output is weighted more heavily since $\alpha_1 \geq \alpha_2$. In fact, if Network 1 makes no mistakes during training, i.e., $c_1^2 = 0$, the weight for the other network α_2 is set to zero so that the output of Network 1 is taken as the ensemble output. When the correlation of the errors by the two networks are low, i.e., when they make mistakes at different input patterns, the ensemble mean square error simpliflifies to

$$\frac{c_1^2 c_2^2}{c_1^2 + c_2^2} \tag{3.17}$$

Suppose Network 1 has a better performance so that $c_2^2 = \rho c_1^2$, where $\rho \geq 1$. The ensemble mean square error is then

$$\frac{\rho}{\rho + 1} c_1^2, \tag{3.18}$$

which is less than the smaller of the mean square errors of the two individual networks. In this case, we can see that the improvement, in the mean square sense, of having an ensemble network can be as high as a factor of 2, corresponding to the case of $\rho = 1$, or when the two networks make equal amounts of mistakes in an uncorrelated fashion.

Consider another scenario. Suppose the two networks make about the same amount of mistakes in the training set and that the mistakes are of approximately the same order. Further, suppose that their mistakes overlap, i.e., they make the same mistakes for a subset of the training set. Tn this case, $c_1^2 = c_2^2$ and $c_{12} = \beta c_1^2$, where β is the amount of overlap of their mistakes, so that $0 \leq \beta \leq 1$. The mean square error of the out put is then

$$\frac{c_1^2}{2}(1 + \beta). \tag{3.19}$$

When the two networks makes the same set of mistakes, $\beta = 1$, and the output mean square error is the same as that of each individual network. When the two networks

make different mistakes, β approaches 0 so that the ensemble output is improved in the mean square sense by a factor approaching 2.

In the general case when $K \geq 2$, a genetic algorithm-based selective neural network ensemble method (GASEN) [27] can be used. A genetic algorithm is used to search for a solution of E_α by defining $h(\alpha) = 1/E_\alpha$ as the fitness function which is to be maximized. The components of a candidate solution α may violate the constraints during its evolution, therefore its elements α_i should be normalized to $\alpha_i / \sum_j \alpha_j$ at each generation.

There are different strategies in using the weights found by the genetic algorithm. In GASEN [27], the weight are used to eliminate networks rather than to be used in a weighted average. We can exclude those networks which are associated worth weights less than a threshold value $\lambda \geq 0$; i.e., exclude Network i if $\alpha_i \leq \lambda$. After those networks are discarded, the outputs of the remaining networks are averaged to form the ensemble output. This is justified by observing that after eliminating those networks that are particularly unreliable, the remaining individual networks have approximately equal performance.

In our work, we set λ to zero and use the weights found by the genetic algorithm to form a weighted average as the ensemble output. Empirically we found this method to be more desirable than using a simple average with or without discarding networks with particularly unrealiable outputs.

3.4 Results

Our preliminary results are published as below:

1. S. Chiewchanwattana, and C. Lursinsap: "FI-GEM Network for Incomplete Time-Series Prediction," *Proceeding International Joint Conference on Neural Network,*

(IJCNN'02) IEEE World Congress on Computational Intelligence, Honolulu, USA, May 12-17, 2002, vol. 2, pp. 1757-1762.

2. S. Chiewchanwattana, C. Lursinsap and C. H. Chu: "Time-Series Data Prediction Based on Reconstruction of Missing Samples and Selective Ensembling of FIR Neural Networks," *Proceedings of The 9th International Conference on Neural Information Processing (ICONIP'02)*, Singapore, November 18-22, 2002, p. 2152 -2156

3. S. Chiewchanwattana, C. Lursinsap and C. H. Chu:"A Reconstructed Missing Data-Finite Impulse Response Selective Ensemble (RMD-FSE) Network," *A Chapter in Neural Information Processing System : Research and Development (Series: Studies in Fuzziness and Soft Computing), vol. 152, Springer, Rajapakse, Jagath C.; Wang, Lipo (Eds.) 2004*, ISBN:3-540-21123-3, p.113-127.

The first and the second problems in section 2, were used to study in our publication 1 [17]. For publication 2 [21], we studied the the second and the third problems. Furthermore, we extended [21] to publication 3 [22]. Our preliminary result are described as follow:

3.4.1 Results of FI-GEM network [17]

Our solution is to develop a new neural network model for forcasting incomplete time-series data and improve the accuracy of prediction. Mackey-Glass chaotic time-series data and the annual sunspots are used to be the two case studies in our preliminary experiment. Various versions of EM-based algorithm and smoothing spline interpolation are used to preprocessing the incomplete data sets. We have to trained the individual networks by supervised neural networks multilayer perceptron(MLP) with extended Kalman filtering [25] by Singhal. The ensemble construction [24]Perrone is used for the

combination of the individual networks. We name this type of network *Fill-In - Generalized Ensemble Method (FI-GEM)*networks. The experimental results are summarized as follows. The missing at random (MAR) incomplete time-series are created by randomly sampling the missing time steps from the training set of the both data sets. The levels of missing values considered in our experiments are set to 2.5%, 5%, 7.5 10%, 12.5%, and 15%.

The performance of prediction by an MLP are evaluated by measuring the difference between the MSE of the desired networks and the reference networks explained in [17].

For Mackey-Glass chaotic time-series data, the input vector for predicting x_t is $[x_{t-6}\ x_{t-12}\ x_{t-18}\ x_{t-24}]^T$. The prediction performances are shown in Figure 3.2(a). We found that some individual networks yield P_d less than zero at some level of missing but they yield P_d greater than zero at another level of missing. In Figure 3.2(a), FI-GEM networks yield P_d less than zero for every level of missing. This shows that FI-GEM networks give the better prediction performance than the reference networks. Furthermore, they yield the lowest P_d also. $P_d\prime$ of six levels of missing: 2.5%, 5%, 7.5% 10%, 12.5%, and 15%, are shown in Table I. The experimental results show that all $P_d\prime$ values are greater than zero. We can notice that the individual networks give the worse prediction performance than FI-GEM networks in terms of both mean square error and consistency.

For the sunspots data, the input vector for predicting x_t is $[x_{t-1}\ x_{t-2}\ \cdots\ x_{t-12}]^T$. The prediction performance are shown in Figure 3.2(b). When we consider the results, we found some individual networks yield P_d less than zero at 7.5% and 10% missing but FI-GEM networks yield P_d less than zero for every level of missing. Furthermore, FI-GEM networks yield the lowest P_d. $P_d\prime$ of six levels of missing: 2.5%, 5%, 7.5% 10%, 12.5%, and 15%, are shown in Table 3.1. All $P_d\prime$ values are also greater than zero. Hence, FI-GEM networks exhibit better performance than the individual neworks.

We have proposed FI-GEM network, which is the incorporation of fill-in techniques and ensemble network for predicting the time-series data in the future. The performance index which measures the accuracy of prediction for the desired network with respect to the individual network using the complete data as the training set is evaluated. Our experiments signify that FI-GEM outperforms each individual network when tested with two most referred benchmarks, Mackey-Glass chaotic time-series and the annual sunspots.

3.4.2 Results of RMD-FSE network [21]

For RMD-FSE network, the ensemble network therefore consists of independently trained neural networks, each drawing an input stream from a fill-in method, which are then combined as a single master network. Each individual network is one that uses a Finite Impulse Response model [26] to perform the prediction. The outputs of all individual neural networks are combined by the genetic algorithm-based selective neural network ensemble method (GASEN) [27]: We denote this approach as a *reconstructed missing data-finite impulse response selective ensemble (RMD-FSE)* network.

Two different data sets are used in our experiment. The first set is the sunspot data and the second set is the daily gauge height. The experimental results of these two data sets are concluded as follows.

For gauge hight data, we choose the first 2,000 days to be the training set and the rest 495 days to be the test set. For sunspots data, we choose the first 245 days to be the training set and the rest 50 years to be the test set. We create the missing at random (MAR) and repeated the mumber of runs like [17].

The performance of prediction by FIR are evaluated by measuring the difference between the MSE of the desired networks and the reference networks. Those are explained

in [17].

For the sunspots data, the weights assigned to each individual network of the ensemble network are shown in Figure 3.3(a). We can see from Figure 3.3(a) that the weights associated with some individual networks at 12.5% and 15% missing input were set to zero by the genetic algorithm. The output of those networks were therefore essentially discarded from the ensembling. The prediction performance is shown in Figure 3.4(a). When we consider the results, we found some individual networks yield P_d less than zero at 2.5%, 5%, 7.5% and 12.5% missing data but RMD-FSE networks yield P_d less than zero for every percentage of missing data. Furthermore, RMD-FSE networks yield the lowest P_d and, consequently, give the better prediction performance than the reference network. The experimental results in Table 2 show that, all P'_d values are also greater than zero. Hence, RMD-FSE networks exhibit better performance than the individual networks.

For the daily gauge height at Ban Luang gauging station, Mae Tun stream, Ping river data, the weights assigned to each individual network of the ensemble network are shown in 3.3(b). Some individual network are discarded from the ensembling at 2.5 %, 7.5 % and 15 % missing. The prediction performances are shown in Figure 3.4(b). The P_d of the RMD-FSE networks are less than zero at 2.5%, 5%, 7.5%, 10% and 12.5% missing data, while at 15% missing data, it is greater than zero. This shows that RMD-FSE networks can work as well as the network which is trained with no missing data. They yield lower P_d than the individual networks at every level of missing data. These results signify that RMD-FSE networks also give better prediction performance than the individual networks. The experimental results in Table 3.2 show that, all P'_d values are greater than zero. We note that the individual networks give worse prediction performance than RMD-FSE networks in terms of both mean square error and consistency.

A novel RMD-FSE network model which is the incorporation of fill-in techniques

and the GA-based selective ensembling of FIR networks is proposed for highly accurate incomplete time-series prediction. In most cases, RMD-FSE network gave the best prediction performance in our experiments. However, when we compare RMD-FSE network and the network which was trained by complete data in the gauge high problem, the difference of prediction performance computed from $(Ensemble - Complete)/Complete \times 100$, is less than 0.5% on average. Hence, in the worst case RMD-FSE network can still work in the same degrees of comparison as the network which was trained from the complete data.

We can conclude that the more correctly estimated missing value, the better prediction accuracy is. To enhance the prediction accuracy of the whole ensemble network, the importance of each individual network must be weighted proportionally to its MSE value.

3.4.3 Results of extended RMD-FSE network [22]

Two different data sets are used in our experiment. The first set is the sunspot data and the second set is the daily gauge height. The prediction results of individual networks and the results of the averaging ensemble network, whose the outputs are averaged from those individual networks, are compared with the results of RMD-FSE network. The experimental results of these two data sets are concluded as follows.

For the sunspot data, the weights assigned to each individual network of the ensemble network are shown in 3.5(a). We can see from Figure 3.5(a) that the weights associated with some individual networks at every percentage of missing input were set to zero by the genetic algorithm. The output of those networks were therefore essentially discarded from the ensembling. The prediction performance is shown in Figure 3.6(a). When we consider the results, we found some individual networks yield P_d less than zero at 2.5%,

7.5%, 10%, 12.5% and 15% missing data but the averaging ensemble network and RMD-FSE networks yield P_d less than zero for every percentage of missing data. Furthermore, RMD-FSE networks yield the lowest P_d and, consequently, give the better prediction performance than both of the reference network and the averaging ensemble network. We note that when the percentage of missing data is high, viz. at 10%, 12.5% and 15%, the network with the average EM selection gives the better result than the network with the random EM selection. The experimental results in Table 1 show that the minimum and maximum P'_d are 1.025×10^{-4} and 8.686×10^{-4}, respectively. Once the complete data are used as the training set of the FIR network, P'_d is 3.330×10^{-4}. All P'_d values are also greater than zero. Hence, RMD-FSE networks exhibit better performance than the individual networks.

For the daily gauge height at Ban Luang gauging station, Mae Tun stream, Ping river, the weights assigned to each individual network of the ensemble network are shown in Figure 3.5(b). Some individual network are discarded from the ensembling at 7.5 %, 10%, 12.5% and 15 % missing. The prediction performances are shown in Figure 3.6(b). Some individual networks yield P_d less than zero at 2.5% and 7.5%missing data. The P_d of the averaging ensemble networks are less than zero at 2.5% and 7.5% missing data, while at 10%, 12.5% and 15% missing data, they are greater than zero. But RMD-FSE networks yield P_d less than zero for every percentage of missing data. Furthermore, RMD-FSE networks yield the lowest P_d and, consequently, give the better prediction performance than both of the reference network and the averaging ensemble network. These results signify that RMD-FSE networks also give better prediction performance than the individual networks. The experimental results in Table 3.3 show that the minimum and maximum P'_d are 1.578×10^{-5} and 22.730×10^{-5}, respectively. When the complete data are used as the training set of the FIR network, P'_d is 1.578×10^{-5}. All P'_d values are greater than zero.

Considering the results of both studies in the experiments, it can be seen that using only one fill-in technique will not achieve desirable performance. Thus, we combine several EM-based estimation methods and spline interpolation for filling in the missing data and ensemble network prediction. RMD-FSE network gave the best prediction performance in our experiments. However, when we compare RMD-FSE network and the averaging ensemble network which was averaged from the individual networks in two problems, the prediction performance of RMD-FSE is better than the averaging ensemble network.

3.5 Summary

Incomplete data sets can be problematic when a neural network is used for time-series prediction. We use a variety of fill-in techniques to generate multiple input streams for an ensemble of MLP or FIR neural network. The new network models, refered to as the FI-GEM and RMD-FSE networks are the ensembling of the outputs of these individual networks. We evaluated the accuracy of prediction with a performance index which measures the accuracy of prediction for the desired network with respect to the individual networks. We conducted our experiments using Mackey-Glass chaotic time-series, the annual sunspot and the daily gauge height data collected at the Ban Luang gauging station, Mae Tun stream, Ping river, Thailand. Our results show that both of FI-GEM and RMD-FSE outperform each individual network, and both of them are proposed for highly accurate incomplete time-series prediction.

3.5.1 Cited Reference

The publication [17] have been used as the applications to the atmospheric sciences [28]. Our work was cited for the reference of the methods for imputation of missing values in

air quality data sets.

3.5.2 Limitation

There are some limitations of both of FI-GEM and RMD-FSE network. First, using an ensemble of networks increase the computational resources needed. Secondly, while the quality of the ensemble output is better than that of the individual networks, it can only be improved by having better individual networks. We have the further work for improving the accuracy of the imputation of incomplete data in the next chapter.

Table 3.1: Average performance index P'_d for Mackey-Glass chaotic time-series and sunspots data. There are six types of individual networks, which are used to compare with FI-GEM networks. We found that all values of P'_d are greater than zero.

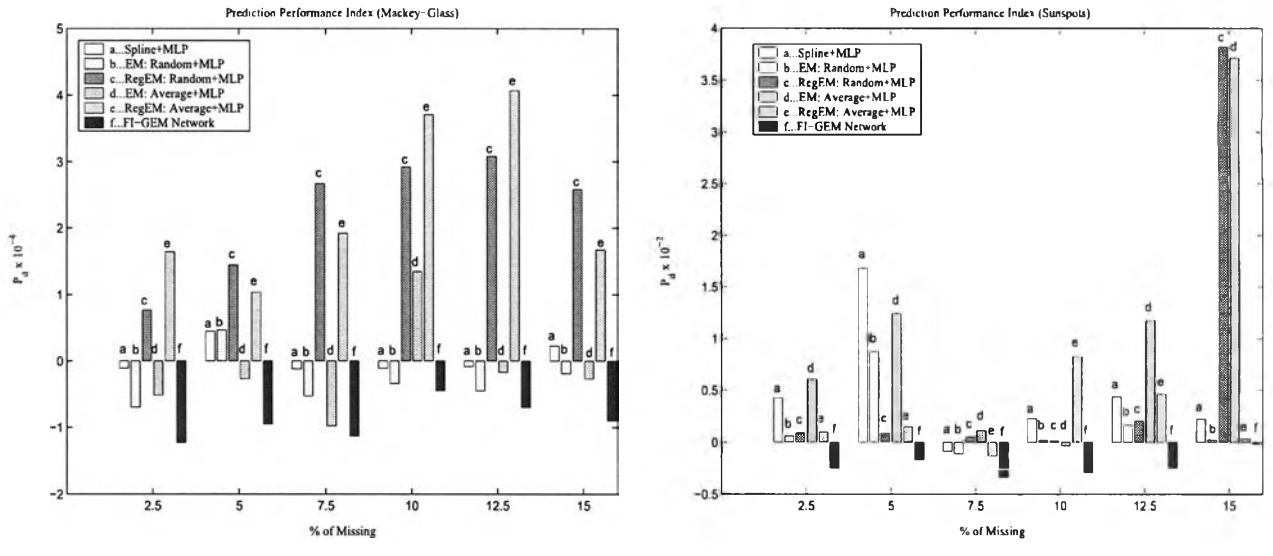
Fill-In Methodology	**Mackey-Glass** $\times 10^{-4}$	**Sunspots** $\times 10^{-2}$
Spline	0.94	0.70
EM(random)	0.60	0.39
Reg EM(random)	3.14	0.93
EM(average)	0.75	1.36
Reg EM(average)	3.24	0.46
Complete Data	0.89	0.22

Table 3.2: Average performance index P'_d for sunspots data and the daily gauge height data. We found that all values of P'_d are greater than zero.

Fill-In Methodology	**Sunspots** $\times 10^{-4}$	**Gauge Height** $\times 10^{-5}$
Spline	6.616	4.825
EM(random)	1.286	8.490
Reg EM(random)	5.499	22.210
EM(average)	7.025	3.893
Reg EM(average)	6.978	19.843
Complete Data	4.715	1.059

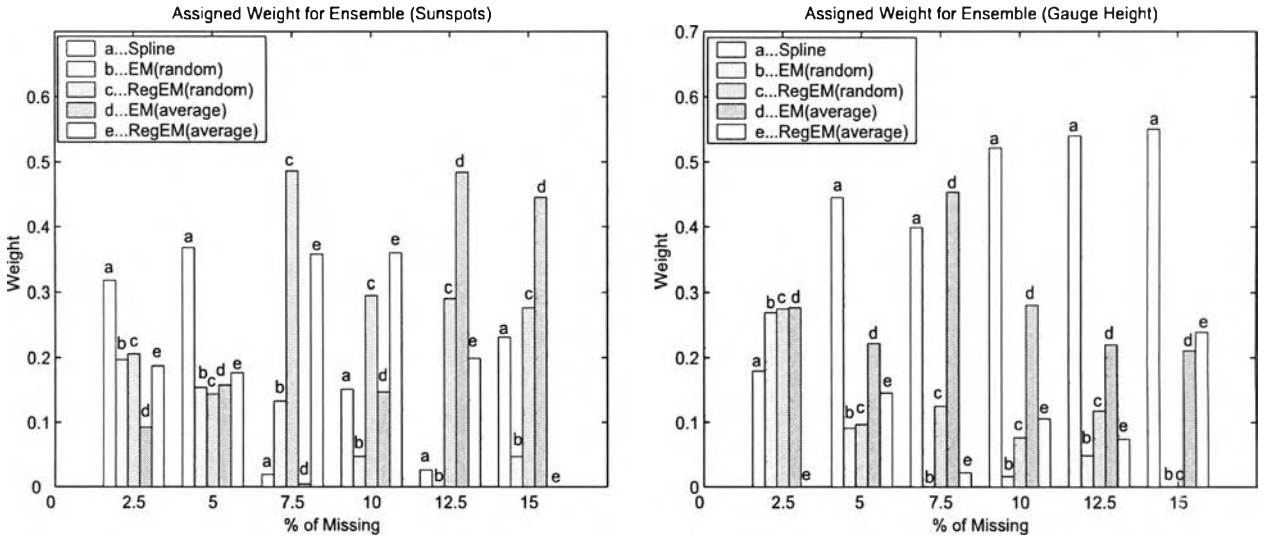
Table 3.3: Average performance index P'_d for sunspots data and the daily gauge height data. We found that all values of P'_d are greater than zero.

Fill-In Methodology	**Sunspots** $\times 10^{-4}$	**Gauge Height** $\times 10^{-5}$
Spline	2.928	5.345
EM(random)	8.686	9.009
Reg EM(random)	4.487	22.730
EM(average)	5.129	4.412
Reg EM(average)	3.076	20.362
Principal Component	5.735	2.505
Averaging Ensemble	1.025	2.226
Complete Data	3.330	1.578



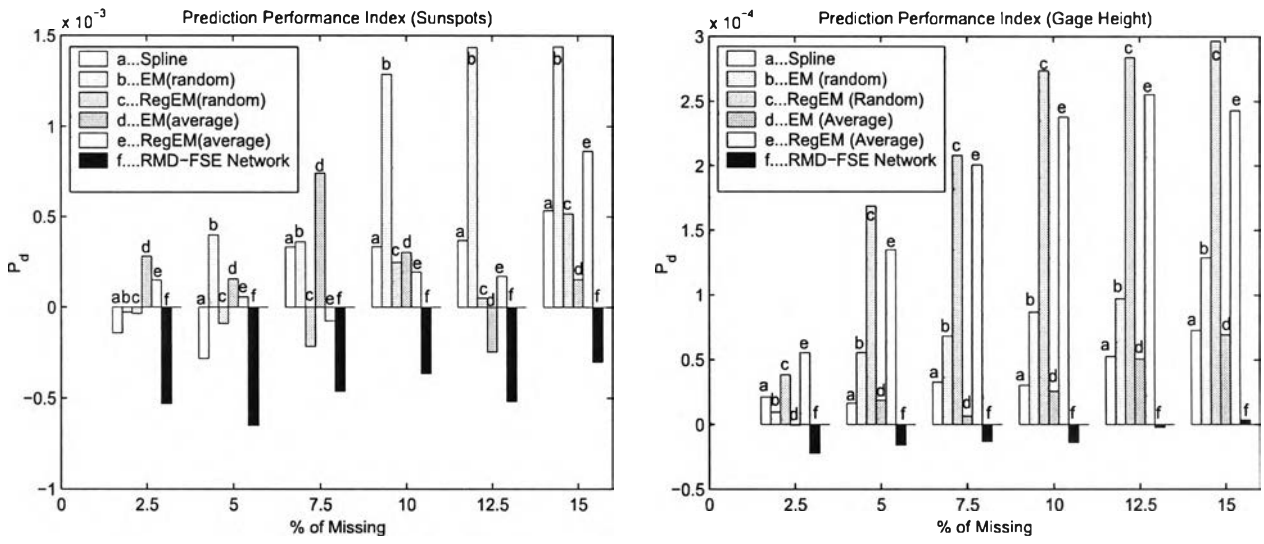
(a) (b)

Figure 3.2: Prediction performance index for (a) Mackey-Glass, and (b) sunspots. There are six groups of bars for 2.5%, 5%, 7.5%, 10%, 12.5% and 15% missing as shown on the x-axis. Each Group has five individual networks and FI-GEM networks. From left to right, the first bar is for MLP for cubic smoothing spline, the second bar is for MLP for EM with random selection, the third bar is for MLP for the regularized EM with random selection, the fourth bar is for MLP for EM with average selection, the fifth bar is for MLP for the regularized EM with average selection, and the last bar is for FI-GEM network. Prediction performance index P_d is shown on the y-axis. The prediction performance of our proposed are the lowest at all levels of missing.



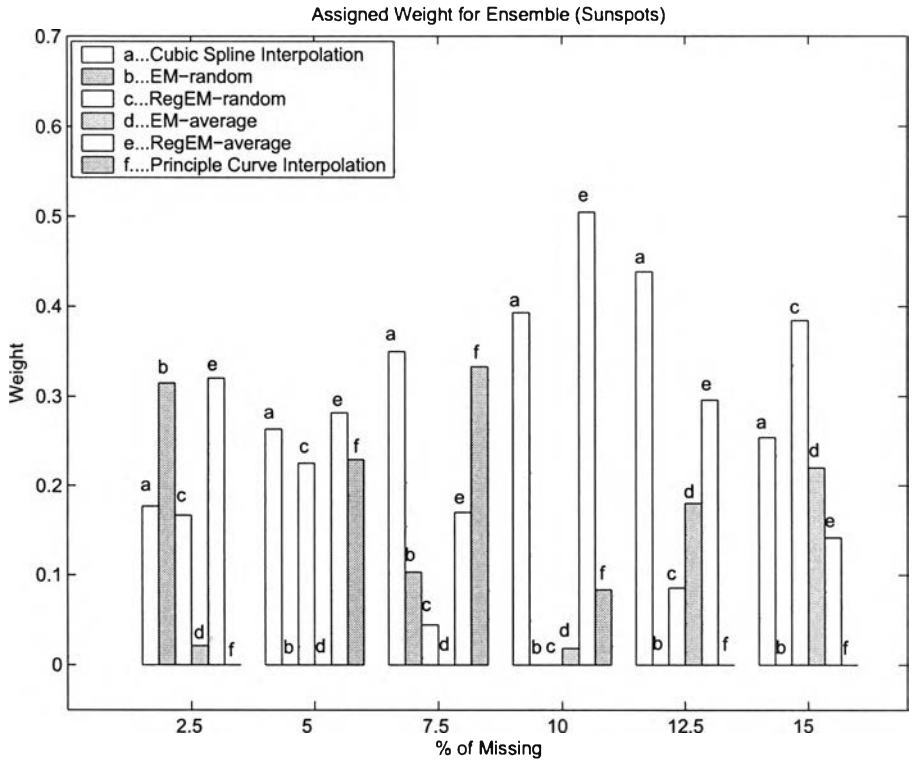
(a) (b)

Figure 3.3: Weight in the selective ensemble network assigned to different FIR networks for the (a) sunspots and (b) gauge height data sets. They differ depending on the amount of data points missing.

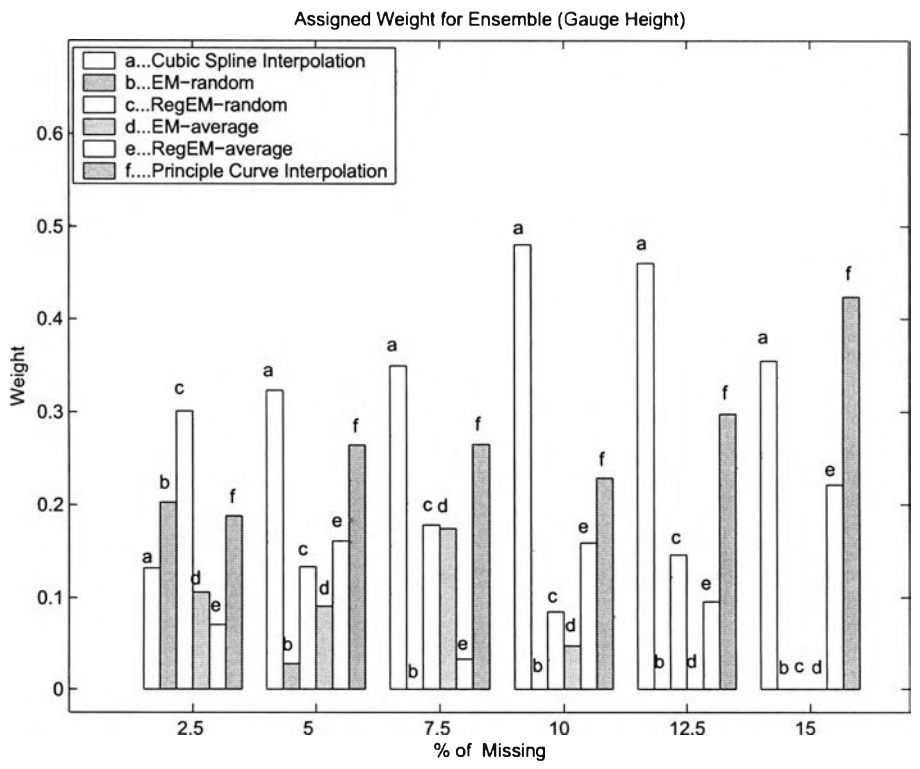


(a) (b)

Figure 3.4: Prediction performance index for the (a) sunspots and (b) gauge height data sets

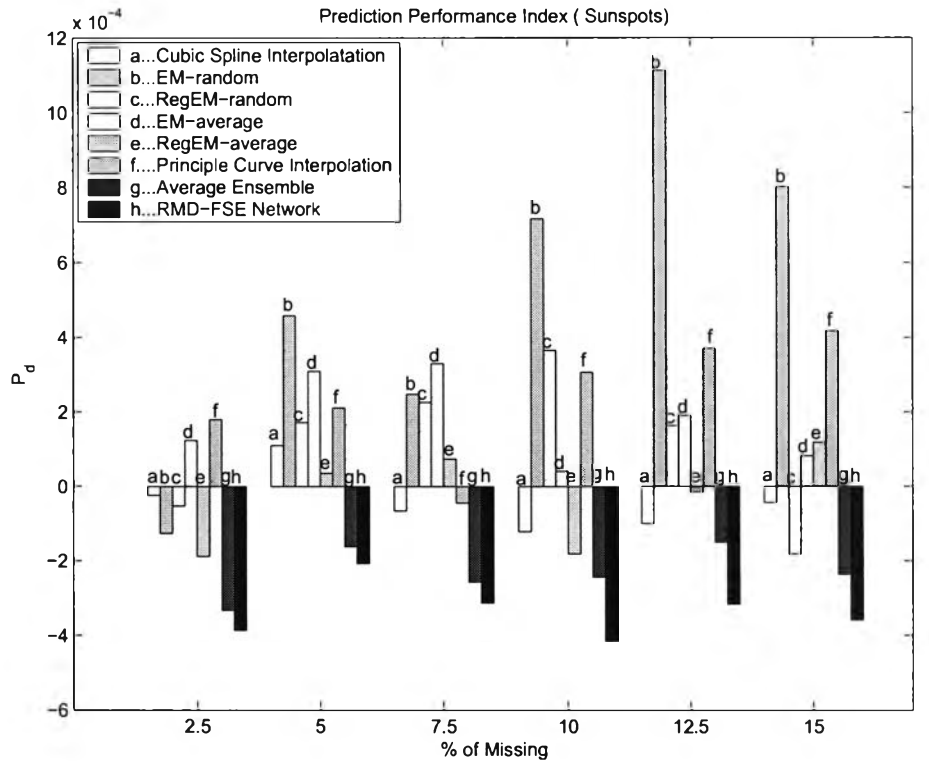


(a)

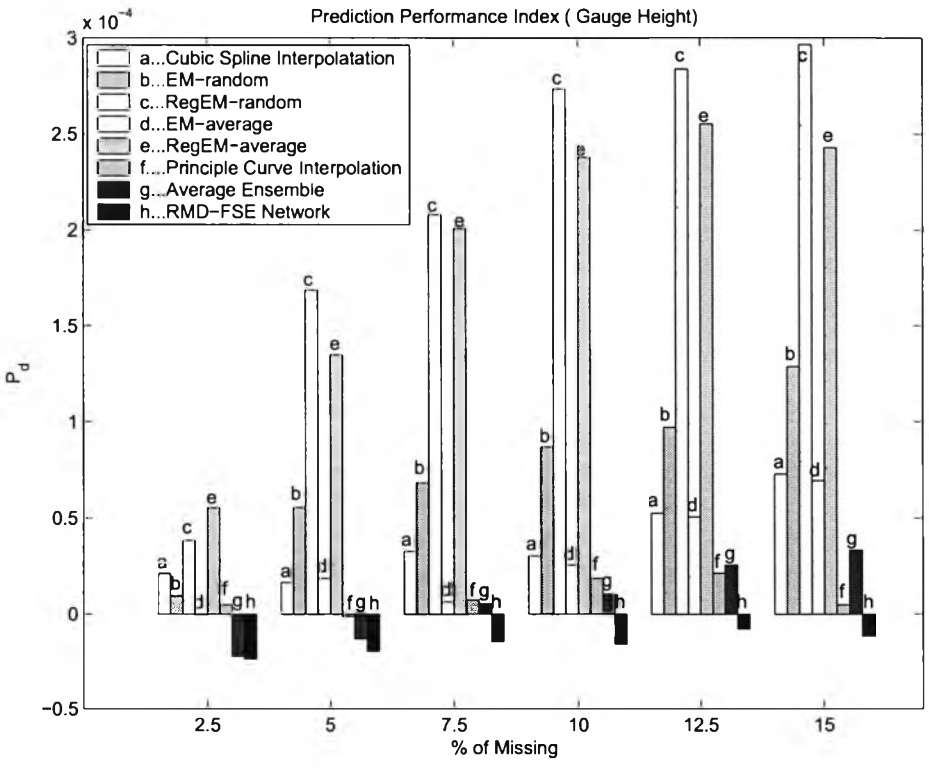


(b)

Figure 3.5: Weight in the selective ensemble network assigned to different FIR networks for the (a) sunspots and (b) gauge height data sets. They differ depending on the amount of data points missing.



(a)



(b)

Figure 3.6: Prediction performance index for the (a) sunspots and (b) gauge height data sets.