Journal of Mathematics and Statistics 9 (4): 349-356, 2013
ISSN: 1549-3644

doi:10.3844/jmssp.2013.349.356 Published Online 9 (4) 2013 (http://www.thescipub.com/jmss.toc)

BAYESIAN MODEL AVERAGING WITH MARKOV CHAIN MONTE CARLO FOR CALIBRATING TEMPERATURE FORECAST FROM COMBINATION OF TIME SERIES MODELS

Heri Kuswanto and Mega Rahmatia Sari

Department of Statistics, Faculty of Mathematics and Natural Science,
Institut Teknologi Sepuluh Nopember (ITS) Surabaya, Indonesia

Received 2013-04-06; Revised 2013-10-03; Accepted 2013-11-29

ABSTRACT

Global warming is an important issue related to the climate and weather forecast. It is shown by significantly increasing the atmospheric temperature level. Hence, improving the forecast accuracy of temperature is an important issue. The forecast is commonly done by performing a deterministic forecast meaning that the system will generate a point forecast without taking into account the uncertainty induced by model specification as well as the nature behavior. Ensemble forecast has been introduced to overcome this problem and it has been implemented in many Ensemble Prediction Systems (EPS) over the world. A problem arises in some developing countries that unable to develop such EPS due to the system restrictions. This paper discusses the performance of combined forecasts generated from a class of time series model as an alternative of EPS. The models are calibrated using Bayesian Model Averaging (BMA) where the parameters are estimated by Markov Chain Monte Carlo (MCMC). The results show that the proposed procedure is capable to increase the reliability of the forecast.

Keywords: Ensemble Prediction System, ARIMA, BMA, Reliable

1. INTRODUCTION

Global Warming is an important issue related to climate and weather change. Global warming has increased the intensity of extreme events towards weather variables leading to climate change. Climate change is defined as the gradually change of air temperature, humidity, atmospheric air pressure, sun and rain intensity as well as wind speed within fifty to hundred years. Therefore, temperature as one of the climate change indicators is important to be forecasted.

There are two kinds of forecasts with regards to the output i.e., deterministic forecast and probabilistic forecast. By deterministic forecast, the system will generate a point forecast, while probabilistic forecast generates an interval forecast representing some degrees of uncertainty. It is well known that future projection of nature behavior such as temperature is highly affected by uncertainty and hence, implementing probabilistic forecast is expected to produce more reliable forecast than deterministic one. Some researches applying deterministic approach for temperature prediction are (Tektas, 2010; Hippert *et al*., 2000; Prybutok *et al*., 2000; Saima *et al*., 2011; Soe *et al*., 2012) among others. Indonesia is one of the countries that currently still implements deterministic forecast to some climate variables.

The probabilistic forecast is a result of using several model outputs for the forecast, that is known as ensemble forecast (Gneiting, 2008; Gneiting *et al*., 2005; Murphy, 1998; Ustaoglu *et al*., 2008; Sloughter *et al*., 2013). The idea of the ensemble forecast is to model the uncertainty induced by several factors such as model specification, nature behavior, initialization and the others. Zhu (2005) discusses the sources of uncertainty comprehensively. In fact, ensemble forecast has been applied in many Ensemble Prediction System (EPS) of developed countries such as National Centers for Environmental Prediction in US, the European Centre for Medium-

Corresponding Author: Heri Kuswanto, Department of Statistics, Faculty of Mathematics and Natural Science,
Institut Teknologi Sepuluh Nopember (ITS) Surabaya, Indonesia

Range Weather Forecasts (ECMWF), the Met-Office UK, Meteo-France, Environment Canada, the Japanese Meteorological Agency, the Bureau of Meteorology Australia, the Korea Meteorological Administration and many others. Forecasting by ensemble has been proven to be able to generate reliable forecast compared to deterministic forecast from a single model.

The model outputs used in the ensemble forecast of EPS involves of numerical processes with relatively high degree of complexity. This leads to a problem in implementation of EPS, in particular for developing countries due to unavailability of resources such as supercomputers. This paper proposes an alternative of numerical ensemble forecast using outputs of several time series models as the ensemble. The ensemble forecast has characteristic of being underdispersive (Raftery *et al*., 2005) and hence it has to be calibrated to remove the bias and match the distribution of observation with forecasts. Several calibration methods have been introduced in many previous researches such as Bayesian Model Averaging (BMA) of (Raftery *et al*., 2005; Wilson *et al*., 2007; Sloughter *et al*., 2013), Dressing kernel of (Wang and Bishop, 2005), Model Output Statistic of (Gneiting and Raftery, 2007; Soe *et al*., 2012) and many others. The BMA method is the most widely used method in the EPS. There are two common procedures used to estimate the parameters of the BMA i.e., Expectation Maximization and Markov Chain Monte Carlo (MCMC), referred hereafter to as BMA_EM and BMA_MCMC respectively. The BMA_MCMC was firstly introduced by (Vrugt *et al*., 2008). The paper showed that the MCMC procedure works well in predicting the wind speed and offers some flexibility in the application. They also pointed out several advantages of using MCMC compared to EM algorithm.

In this study, the BMA_MCMC is applied to ensemble temperature forecast where the outputs are generated from several time series models instead of generated from numerical process. This approach can be considered as an extension of forecast combination proposed by (Granger, 1989), where the forecast is combined and calibrated in this study. In this case, we intend to model uncertainty induced by the model specification. The performance of the proposed methodology is compared to the results of BMA_EM documented in (Kuswanto, 2011).

1.1. Literature Review

1.1.1. Bayesian Model Averaging

Bayesian Model Averaging (BMA) is one of the statistical methods for calibration that combine some information from several model outputs. The BMA for dynamic ensemble forecast can be written as:

$$p(y|f_1 ...f_k) = \sum_{k=1}^{K} w_k g_k(\Delta | f_k) \quad (1)$$

where, w_k is the positive posterior probability of k-th forecast with sum of one and K is the number of ensemble members.

Equation (1) requires a specification about the prior distribution of the underlying climate variable (e.g., temperature) which is approximated by normal distribution such that:

$$y | f_k \sim N(a_k + b_k f_k, \sigma^2)$$

where, a_k and b_k are the bias correctors derived from simple linear regression of y to f_k for every ensemble member. The mean of the BMA forecast is given by Equation (2):

$$E\{\Delta | f_1 ...f_k\} = \sum_{k=1}^{K} w_k (a_k + b_k f_k) \quad (2)$$

while the variance of the BMA at period t is:

$$var[y_t | f_{1t},...,f_{Kt}] = \sum_{k=1}^{K} w_k((a_k + b_k f_k) - \sum_{l=1}^{K} w_l(a_l + b_l f_l)) + \sigma^2$$

1.2. Markov Chain Monte Carlo (MCMC)

The BMA parameters i.e., variance and weight are estimated using Sampling of MCMC. The MCMC simulation allow us to estimate parameters of a complex posterior distribution with high dimensionality. Granger (1989) proposes an algorithm called as Differential Evolution Adaptive Metropolis (DREAM) where N different Markov chains are run simultaneously in parallel. If the state of a single chain is given by a vector θ with dimension of d, thus each generation of N in DREAM defines a population Ω with dimension N×d. The jump of every chain is generated by randomly taking the difference among several other chains of Ω (without replacement) such as:

$$\vartheta^i = \theta^i + \gamma(\delta)\sum_{j=1}^{\delta} \theta^{r(j)} - \gamma(\delta)\sum_{n=1}^{\delta} \theta^{r(n)} + e$$

where, δ represents the number of pairs used to produce the candidates. The ratio metropolis is used as the criteria to decide whether to accept or reject the candidate.

1.3. Continuous Ranked Probability Score (CRPS)

The CRPS measures how reliable the calibration result of probabilistic forecast. The formula for calculating the CRPS is given by:

$$\text{CRPS} = \frac{1}{K}\sum_{i=1}^{K}\int_{x=-\infty}^{x=\infty}(F_i^f(x) - F_i^0(x))^2 dx$$

where, $F_i^f(x)$ is cumulative density function (cdf) of the i-th forecast, while $F_i^0(x)$ is the cdf of true observation and K is the number of ensemble member. Small value of CRPS indicates that the forecast is reliable (Gneiting and Raftery, 2007).

2. MATERIALS AND METHODS

2.1. Data

This study analyzes daily mean temperature observed at climatological Station Juanda-Indonesia spanning from 2008-2009. The data is divided into two parts i.e., training and testing. The training data is used to generate build the time series model for generating the ensemble, while the testing data is used for evaluation of the forecast performance. We use different sizes of training windows to implement the BMA i.e., 10, 15, 20 and 25.

2.2. Steps of the Analysis

The steps of the analysis that is carried out in this study are described as follows:

- Plot ensemble for verification of the generated outputs
- Correct the bias of the mean by performing regression between true observation with each forecast, where the length of data equals to the training window
- Estimate the parameters of BMA by MCMC using DREAM algorithm. The parameter estimation is carried by following procedures:
  - Determine the training windows
  - Use sampling of 50000 for the parameter space
  - Determine the number of Markov chain
  - Omit the outlier chain using Inter Quartile Range (IQR)
  - Estimate the parameters by Maximum likelihood
- Construct the mixture distribution
- Calculate the calibrated mean and variance
- Generate 1000 data following the mixture distribution as the predictive distribution
- Evaluate the calibration results and compare it to the raw ensemble

3. RESULTS AND DISCUSSION

The mean daily temperature observed at meteorological station Juanda Indonesia is 27,63 degree with the maximum reached 32.6 degree and minimum of 24.2 degree. The distribution of the data is approximately normal shown by the skewness of 0.32. It validates the previous assumption that the temperature in this case is approximated by normal distribution as prior distribution. In this study four time series models are constructed and the forecasts generated from these models will be combined or calibrated using BMA. The models belong to the class of ARIMA. The selected models are ARMA(2,1), ARIMA(1,1,1), ARIMA(1,1,2), IMA(0,1,3) denoted hereafter to as M1, M2, M3 and M4 respectively. These models are the best models chosen among several candidate models by considering some rules in the ARIMA modeling.

Figure 1 and 2 shows the time series plots of the 1 day and 7 day lead forecasts.

It is clear that the spread of the forecasts is underdispersive meaning that the have a low spread and tend to generate similar values for all models. The 7 day forecast yield on more bias than 1 day forecast. It is therefore necessary to calibrate the forecasts and we apply BMA_MCMC as the calibration method. The illustration of the bias correction as the result of regression between ensemble member and observation is given in **Table 1**. The mean in the last column shows the deterministic forecast for each member that will be combined with some degrees of model performace represented by a weight for each. Meanwhile **Table 2** illustrates the bias correction for forecast on 26th December 2009.

Note that the observation on 29th December is 28 degree which means that bias correction on that date is capable to generate forecast with low bias. The parameters estimated by MCMC procedure using different training windows are shown in **Table 2**.

The BMA combines M1 to M4 with respects to its performance (represented by the weight) and yields on the following probabilistic forecast. We perform the corresponding CRPS for each training window.

From **Table 3**, we know that the minimum CRPS is 0.2494 obtained using 10 days training window and this is the optimum setting. **Figure 3** depicts the predictive distribution of the calibrated forecast using training window of 15 days.

The predictive distribution in **Fig. 3** clearly shows the performance of each model in the forecast. Higher weight represents better the performance of the model in the system. We replicate the similar procedure to calibrate the 7 day forecast for the forecast on 6th December 2009. **Table 4-6** show the parameters of the BMA i.e., bias correction, BMA parameters and the predictive distributions respectively.

We observe different result of the predictive distribution in term of the optimum training window. However, the bias as well as the interval of the forecast are nearly the same as 1 day forecast. The optimum setting is obtained using 20 day training window.

The performance of the calibrated ensemble forecast is evaluated from the system. It means that the choice of the optimum training window is based on the performance of each setting for the whole periods of validation. The criteria of the assessment is CRPS. The CRPS value reflects the reliability of the forecast. The analysis shows that 1 day forecast has the best performance under 25 days training window while for the best performance for 7 day forecast is shown by 15 training window. The summary of the value is given by **Table 7**. Indeed, CRPS measure the reliability with respect to the bias and compactness of the forecast.

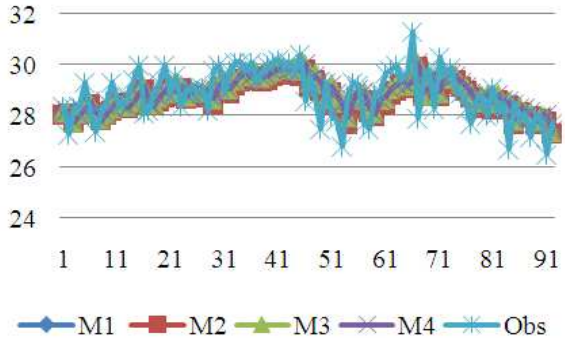

Fig. 1. Plot ensemble for 1 day lead forecast

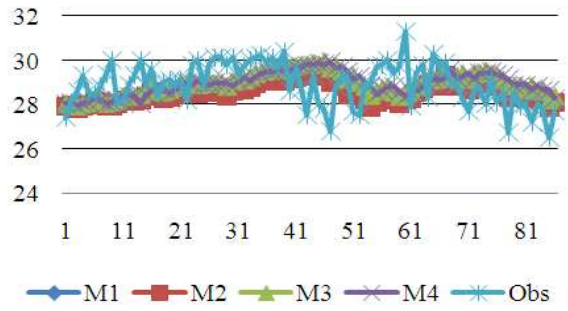

Fig. 2. Plot ensemble for 7 day lead forecast

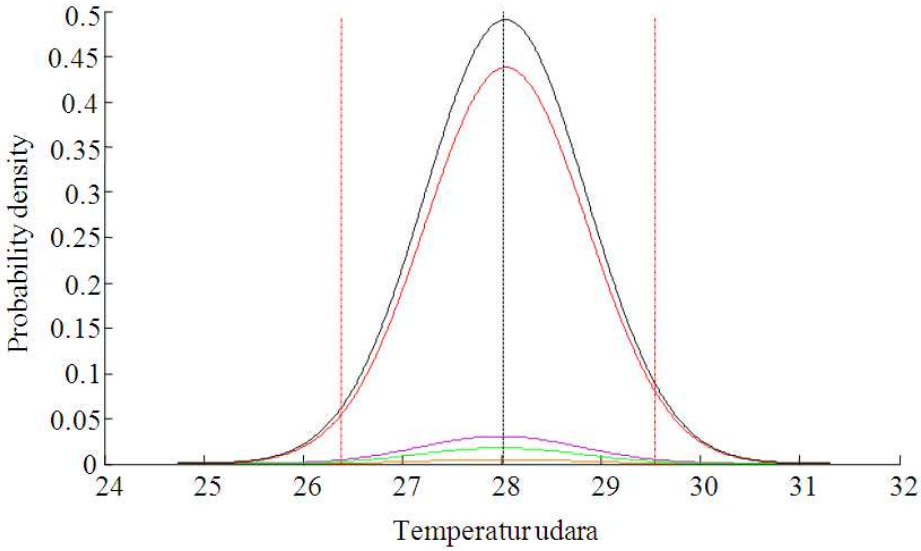

Fig. 3. Predictive distribution for 26th December 2009 using 15 days training window

Table 1. Parameters of bias correction for 1 day forecast

m	Model	a	b	*f*	Mean ($a_{k+}b_kf_k$)
10	M1	44,027	-0,558	28,03	28,399
	M2	45,901	-0,624	27,98	28,434
	M3	41,51	-0,466	28,17	28,380
	M4	37,248	-0,317	28,31	28,276
15	M1	6,681	0,761	28,030	27,999
	M2	6,980	0,751	27,980	27,988
	M3	6,977	0,746	28,170	27,988
	M4	4,365	0,836	28,310	28,045
20	M1	16,983	0,405	28,030	28,343
	M2	16,312	0,429	27,980	28,310
	M3	17,721	0,378	28,170	28,359
	M4	17,202	0,396	28,310	28,400
25	M1	9,531	0,672	28,030	28,367
	M2	8,808	0,697	27,980	28,320
	M3	10,921	0,621	28,170	28,406
	M4	10,429	0,638	28,310	28,479

Table 2. Parameters generated from BMA_MCMC

		M			
Parameter	Model	10	15	20	25
Corrected mean	M1	28,399	27,999	28,343	28,367
	M2	28,434	27,988	28,310	28,320
	M3	28,380	27,988	28,359	28,406
	M4	28,276	28,045	28,400	28,479
Variance	M1	0,608	0,661	0,812	0,917
	M2	0,608	0,661	0,812	0,917
	M3	0,608	0,661	0,812	0,917
	M4	0,608	0,661	0,812	0,917
Weight	M1	0,044	0,010	0,013	0,022
	M2	0,925	0,062	0,919	0,936
	M3	0,008	0,036	0,003	0,039
	M4	0,024	0,892	0,065	0,004

Table 3. Parameters and CRPS comparison for different training lengths (1 day)

m	Mean	Varians	Batas Bawah	Batas Atas	CRPS
10	28,42843	0,608308	26,89974	29,95711	0,2494
15	28,03868	0,613548	26,50342	29,57393	0,3737
20	28,31596	0,60553	26,73508	29,89683	1,498
25	28,32492	0,628511	26,77105	29,87878	0,3437

Table 4. Parameters of bias correction for 7 day forecast

M	Model	a	b	f	Mean ($a_x + b_x f_x$)
10	M1	-11,610	1,384	28,290	27,549
	M2	-9,489	1,314	28,220	27,584
	M3	-11,258	1,352	28, 660	27,482
	M4	6,374	0,746	28,890	27,927
15	M1	-27,970	1,959	28,290	27,456
	M2	-25,491	1,877	28,220	27,472
	M3	-23,657	1,784	28,660	27,481
	M4	14,307	0,485	28,890	28,331
20	M1	-11,490	1,396	28,290	27,989
	M2	-14,842	1,514	28,220	27,892
	M3	4,498	0,829	28,660	28,255
	M4	33,900	-0,180	28,890	28,710
25	M1	41,517	-0,439	28,290	29,108
	M	37,083	-0,285	28,220	29,054
	M3	48,979	-0,690	28,660	29,204
	M4	66,518	-1,294	28,890	29,137

Table 5. Parameters generated from BMA_MCMC

		m			
Parameter	Model	10	15	20	25
Mean terkoreksi	M1	27,549	27,456	27,989	29,108
	M2	27,584	27,472	27,892	29,054
	M3	27,482	27,481	28,255	29,204
	M4	27,927	28,331	28,710	29,137
Varians	M1	0,535	0,567	0,754	0,903
	M2	0,535	0,567	0,754	0,903
	M3	0,535	0,567	0,754	0,903
	M4	0,535	0,567	0,754	0,903
Bobot (*Weight)*	M1	0,041	0,005	0,037	0,022
	M2	0,942	0,967	0,926	0,033
	M3	0,015	0,011	0,003	0,005
	M4	0,001	0,017	0,033	0,940

Table 6. Parameters and CRPS comparison for different training length (7 day)

m	Mean	Varians	Batas Bawah	Batas Atas	CRPS
10	27,58131	0,694269	25,94818	29,21444	0,3125
15	27,48735	0,680943	25,86997	29,10473	0,4068
20	27,9236	0,618088	26,38268	29,46452	0,1827
25	29,13424	0,616912	27,59478	30,67369	0,5072

Table 7. CRPS comparison between uncalibrated and calibrated forecast

	m (Training window)				
Lead	Ensemble	m = 10	m = 15	m = 20	m = 25
1	Uncalibrated	0,6723	0,6814	0,6794	0,6967
	Calibrated	0,5296	0,5579	0,5332	0,5143
7	uncalibrated	0,8948	0,9012	0,9285	0,9501
	calibrated	0,5509	0,5290	0,5544	0,5714
Lead	Method	m = 10	m = 15	m = 20	m = 25
1	BMA-MCMC	0,5296	0,5579	0,5332	0,5143
	BMA-EM	0,535	0,566	0,529	0,510
7	BMA-MCMC	0,5509	0,5290	0,5544	0,5714
	BMA-EM	0,495	0,529	0,544	0,584

In term of the proportion of observation captured by the produced interval, it can be seen in **Fig. 4 and 5**.

From the figures, it can be seen that using 1 day forecast 91% of the observations can be covered by the produced interval, while 7 day forecast is able to capture the observations with the proportion of 92%. It is rational to have this as the widht of interval forecast becomes wider for longer lead time.

The table indicates that the calibration using BMA_MCMC is capable to generate more reliable forecast than the uncalibrated one. It is shown by the lower value of CRPS of calibrated forecast. It means that the calibration will generate more reliable probabilistic forecast.

Having compared the performance of the BMA_MCMC with the unclaibrated forecast, we will compare the performance of BMA_MCMC with BMA_EM. As previously mentioned, the parameters of the BMA can be esimated using those two algorithms. We directly compare the BMA_MCMC with the performance of BAM_EM documented in (Kuswanto, 2011). The study discusses the performance of BMA_EM using the same dataset and ARIMA models as used in this study.

For the optimum training length, both BMA algorithms yield on the same suggestion i.e., using 25 day training window for 1 day ahead forecast. However, different results are observed for lead time of 7 day forecast. BMA_MCMC suggests to use 15 training window while BMA_EM found 10 day as th eoptimum one. Nevertheless, the performance of both algorithms is the same in particular short term forecast. For long lead time forecast, we suggest to use 10 day training window to generate more reliable forecast.

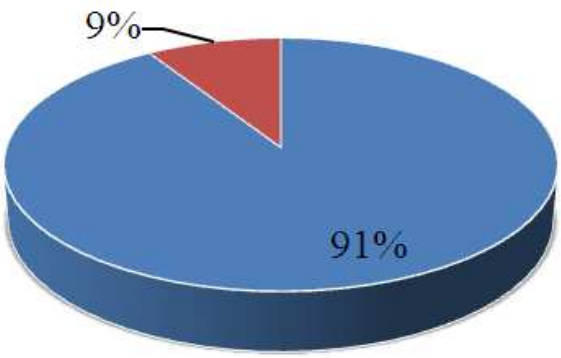

Fig. 4. Proportion of observations captured by the predictive pdf for 1 day lead forecast

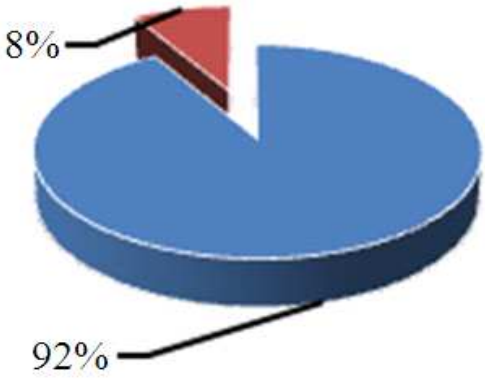

Fig. 5. Proportion of observations captured by the predictive pdf for 7 day lead forecast

4. CONCLUSION

Calibration of the temperature forecast is done by generating four simple time series models which is treated as ensemble members as commonly used in any Ensemble Prediction System. The forecasts are under dispersive and hence it has to be calibrated. By calibration, the bias of the mean forecast is removed and it is expected that the generated probabilistic forecast yields on reliable forecast. Applying BMA_MCMC to calibrate the ensemble forecasts generated from four ARIMA models is capable to produce more reliable forecasts than the uncalibrated forecast. It has been shown also that using the proposed simple procedure works well for short and modest lead time forecasts. Different setting of training windows has been investigated during the implementation of the procedure in order to find the optimum one. Compared to the performance of another BMA algorithm, the BMA_MCMC is unable to outperform BMA_EM. However, both have similar performance and there is gain in using these approaches. Further investigation is necessarily to do in order to study the performance of the BMA applied to models from different classes of time series models such as Transfer function, ANFIS. Moreover, applying BMA to other climate variables generated by similar procedure as applied in this study is worthy to be carried out. Overall, we have shown that using combination of time series models can be a proxy of ensemble forecasts generated from numerical ensemble prediction system. It is very simple and easy to be implemented.

5. REFERENCES

Gneiting, T. and A.E. Raftery, 2007. Strictly proper scoring rules, prediction and estimation. J. Am. Stat. Assoc., 102: 359-378. DOI: 10.1198/016214506000001437

Gneiting, T., A.E. Raftery, A.H. Westveld and T. Goldman, 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Rev., 133: 1098-1118. DOI: 10.1175/MWR2904.1

Gneiting, T.E., 2008. Probabilistic forecasting. J. Royal Statist. Soc. Series A.

Granger, C.W.J., 1989. Combining forecasts: Twenty years later. J. Forecast.

Hippert, H.S., C.E. Pedreira and R.C. Souza, 2000. Combining neural networks and ARIMA models for hourly temperature forecast. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Jul. 24-27, IEEE Xplore Press, Como, pp: 414-419. DOI: 10.1109/IJCNN.2000.860807

Kuswanto, H., 2011. Artificial ensemble forecasts: A new perspective of weather forecast in Indonesia. Proceedings of the International Conference on Mathematics and its Applications, Jul. 12-15, ITS Community.

Murphy, A.H., 1998. The early history of probability forecasts: Some extensions and clarifications. Weather Forecast., 13: 5-15.

Prybutok, V.R., J. Yi and D. Mitchell, 2000. Comparison of neural network models with ARIMA and regression models for prediction of houston's daily maximum ozone concentrations. Eur. J. Operat. Res., 122: 31-40. DOI: 10.1016/S0377-2217(99)00069-7

Raftery, A.E., T. Gneiting, F. Balabdaoul and M. dan Polakowski, 2005. Using bayesian model averaging to calibrate forecast ensembles. Monthly Weather Rev., 133: 1155-1174. DOI: 10.1175/MWR2906.1

Saima, H., J. Jaffar, S. Belhaouari and T. Jillani, 2011. ARIMA based Interval type-2 fuzzy model for forecasting. Int. J. Comput. Applic., 28: 17-21.

Sloughter, M.J., T. Gneiting and A.E. Raftery, 2013. Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. Mon. Wea. Rev., 141: 2107-2119. DOI: 10.1175/MWR-D-12-00002.1

Soe, K., B. Hyeon, S. Hyun and Y. Lee, 2012. Genetic Programming-Based Model Output Statistics for Short-Range Temperature Prediction. In: Applications of Evolutionary Computation, Chio, C.D., A. Agapitos, S. Cagnoni, C. Cotta and F.F. De Vega, (Eds.), Springer, New York, ISBN-10: 3642291775, pp: 122-131.

Tektas, M., 2010. Weather forecasting using ANFIS and ARIMA models. Environ. Res. Eng. Manage.

Ustaoglu, B., H.K. Cigizoglu and M. Karaca, 2008. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorol. Applic., 15: 431-445. DOI: 10.1002/met.83

Vrugt, J.A., C.G.H. Diks and M.P. Clark, 2008. Ensemble bayesian model averaging using markov chain monte carlo sampling. Environ. Fluid Mechan., 8: 579-595. DOI: 10.1007/s10652-008-9106-3

Wang, X. and C. Bishop, 2005. Improvement of ensemble reliability with a new dressing kernel. Q. J. Royal Meteorol. Soc., 131: 965-986. DOI: 10.1256/qj.04.120

Wilson, L.J., S. Beauregard, A.E. Raftery and R. Verret, 2007. Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian model averaging. Monthly Weather Rev., 135: 1364-1385. DOI: 10.1175/MWR3347.1

Zhu, Y., 2005. Ensemble forecast: A new approach to uncertainty and predictability. Adv. Atmospheric Sci., 22: 781-788. DOI: 10.1007/BF02918678