MAUSAM, **74**, 4 (October 2023), 1065-1072

MAUSAM

DOI : https://doi.org/10.54302/mausam.v74i4.4359
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No. 551.577 : 551.509.313.4

New method of precipitation forecast model and validation

KUMARASWAMY KANDUKURI* and BHATRACHARYULU N. CH.

**Kaloji Narayana Rao University of Health Sciences, Warangal,*

Osmania University, Hyderabad, India

(*Received 23 November 2021, Accepted 9 March 2023*)

e mails : kumaraswami.kandukuri@gmail.com; dwarakbhat@rediffmail.com

सार – **पिछले कुछ वर्षों में विभिन्न पूर्वानुमान तकनीकों के साथ कई वास्तविक क्षेत्रों के बहुत सारे समय श्रृंखला डेटा मौजूद हैं। हालाँकि, पूर्वानुमान तकनीकों जैसे कि व्यक्तिगत पूर्वानुमान मॉडल ऑटोरिग्रेसिव, मूविंग एवरेज, ऑटोरिग्रेसिव मूविंग एवरेज, ऑटोरिग्रेसिव इंटीग्रेटेड मूविंग एवरेज, आर्टिफिशियल न्यूरल नेटवर्क, लॉन्ग शॉर्ट टर्म मेमोरी नेटवर्क और ऑटो-रिग्रेसिव कंडीशनल हेटेरोसेडास्टिसिटी / सामान्यीकृत ऑटोरिग्रेसिव कंडीशनल हेटेरोसेडास्टिसिटी और पूर्वानुमान के संयोजन (पूर्वानुमानों का सरल औसत, न्यूनतम भिन्नता विधि, और संयोजन की प्रतिगमन विधि) पर कोई सर्वसम्मत निष्कर्ष नहीं है। अधिकांश आनुभविक जल विज्ञान संबंधी समय श्रृंखला मॉडल मौसम का सटीक पूर्वानुमान नहीं करते हैं। यह शोध पत्र प्रस्तावित हाइब्रिड स्टोचैस्टिक मॉडल (एचएसएम) पूर्वानुमान प्रक्रिया के साथ विभिन्न मौजूदा व्यक्तिगत और संयोजन पूर्वानुमानों के तुलनात्मक अध्ययन पर केंद्रित है। प्रस्तावित पूर्वानुमान मॉडल का परीक्षण करने के लिए हम भारतीय उपमहाद्वीप के जल विज्ञान संबंधी समय श्रृंखला डेटा पर विचार करते हैं। समग्र रूप से अन्य सभी पारंपरिक मॉडल के योगदान सटीकता की तुलना में, प्रस्तावित मॉडल ने अच्छा प्रदर्शन किया और सर्वोत्तम पूर्वानुमान प्राप्त करने के लिए मॉडल में शामिल की जाने वाली पूर्वानुमान तकनीकों की एक इष्टतम संख्या चुनने के लिए हमने मॉडल के आयाम के न्यूनीकरण दृष्टिकोण की भी जांच की।**

ABSTRACT. There is a lot of time series data in many realistic sectors with different forecast techniques over the years. However, there is no unanimous conclusion on forecast techniques such as individual forecast models Autoregressive, Moving Averages, Autoregressive Moving Averages, Autoregressive Integrated Moving Averages, Artificial Neural Network, Long Short Term Memory Network and Auto-Regressive Conditional Heteroscedasticity / Generalized Autoregressive Conditional Heteroscedasticity and combination of forecast (Simple Average of forecasts, Minimum Variance Method and Regression Method of the combine). The most empirical hydrological time series models do not accurately forecast the weather. This paper focuses on a comparative study of different existing individual and combination forecasts with the proposed Hybrid Stochastic Model (HSM) forecast procedure. For this we consider a hydrological time series data of the Indian subcontinent to test the proposed forecast model. As a whole in comparison to all other traditional model's contributions accuracy, the proposed model performed well and also we examined the model's dimension reduction approach to choose an optimum number of forecast techniques to be included in the model to yield the best forecasts.

Key words – ARIMA, Hybrid Stochastic Model, Linear Weighted Average Model, Performance measure, Rainfall.

1. Introduction

There is a lot of time series data in many realistic sectors with different forecast techniques over the years. These are employed in signal processing, pattern recognition, mathematical finance, weather forecasting, engineering control and pretty much any other subject of applied sciences. On a global scale, numerous attempts have been undertaken by various researchers to reliably forecast the weather using various methodologies. However, because of the nonlinear nature of weather, the prediction accuracy achieved by these systems is still insufficient. There is no unanimous conclusion on forecast techniques such as individual forecasts [AR (p), MA (q), ARIMA, ANN, LSTM and ARCH/GARCH, etc.] and combination of forecast (Average of forecasts, Minimum variance method and Regression method of the combine). In order to estimate projected rainfall prediction models with volatility, Kumaraswamy (2022) investigated the precipitation trend pattern throughout India. Forecasting is an uncertain process and there are a variety of prediction approaches and models that integrate these different projections, a method is known as the combination of forecasts, to reduce that uncertainty (Martins & Werner, 2012). Combination approaches have been thoroughly compared since Bates and Granger (1969). Several authors (Bates & Granger, 1969; Clemen, 1989; Krishnamurti *et al*., 2000; Poncela *et al*., 2011, Martins & Werner, 2012 and Kumaraswamy & Bhatracharyulu, 2019 & 2023) have concluded that the combination of forecasts is more accurate than the best individual model. The combination is not enough, though; we also need to know the techniques to utilize and how to combine them. We initially examine the ARIMA, SARIMA and ARCH/GARCH models, which integrate robust methodology for processing stationery and non-stationary time series. In combined regression models, most regression methods are used such as backward, forward and stepwise in variable selection, which, together algorithms in practice to choose variables. All of these algorithms use significance as the sole criterion for included variables in a model.

The comparison is made directly with accuracy measures because this work presents a comparative analysis of accuracy in different forecast techniques. The Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Trend and Theil's U coefficient are the metrics employed. This paper is divided into four sections, the first of which is this introduction. Section two provides a brief on the methodologies used. The comparison findings are presented in section three. Finally, in part four, the study's key conclusions are presented.

2. Methodology

Forecasting is crucial in all sectors of applied sciences and business. Given the devastation that a well-defined structured forecast may bring, more accurate forecasts are desired. For example, when a company's production falls behind current demand, or when a high production leads to greater storage costs and product depreciation, more accurate forecasts are sought. As a result, the concept of a combined prediction emerged, backed up by the promise of a better prognosis than the best individual forecast available.

The combined models are compared to classic time series models on a relative basis. As a whole, in comparison to all other models contributions in terms of accuracy are compared. We proposed new construction of methods for predictions and compared their efficiency with the existed prediction techniques. These techniques are aggregating all the models into one model and also we predict the most accurate results with the reduction of dropping suitable models from the model. Stepwise regression seeks to build a model step by step, only adding or removing predictors that have statistical significance. A single regression model emerges as a result of this procedure. Stepwise analysis can be performed in either a forward or backward direction which the model's dimension reduction approaches to choose an optimum number of forecast techniques to be included in the model to yield the best forecasts.

Here, we presented a new model with other 3 combined models for comparison and named as (*i*) Simple Average model, (*ii*) Weighted Average model, (*iii*) Linear Ensemble model and, (*iv*) Hybrid Stochastic model. The later constructed model is complex than the aforesaid models and each model is described below.

2.1. *Simple Average Model* (*SAM*) : The forecast is the mean of all predicted models and lets there are f_1, f_2, …, f_N models are used for prediction. The SAM is

$$\hat{y}_i = \frac{1}{N}\sum_{t=1}^{N} f_{ti}; t = 1,2..., N; i = 1,2,...m$$

where, f_{ti} = the prediction of t^{th} model for i^{th} observation.

2.2. *Weighted Average Model* (*WAM*) : This method is constructed on aggregation weights of the prediction of the models. The weights are assigned to the models based on their performance criteria. The WAM is

$$\hat{y}_i = \sum_{t=1}^{N} w_t . f_{ti}; t = 1,2..., N; i = 1,2,...m$$

where, w_t = the weight of the t^{th} forecasted model, f_{ti} = the prediction of t^{th} model for i^{th} observation.

2.3. *Linear Ensemble Model* (*LEM*) : This model includes the statistics of the original observations which are used for prediction and as well as the forecasted models. This model is an ensemble of nature and designed as

$$\hat{y}_i = \overline{O_i} + \sum_{t=1}^{N} w_t .\left(f_{ti} - \bar{f}_i\right); t = 1,2..., N; i = 1,2,...m$$

where

$\overline{O_i}$ = the observed mean for the i^{th} observation,

w_t = the weight of the t^{th} forecasted model,

TABLE 1

The rainfall predictions of the fitted traditional models

Time	Observed	ARIMA (2, 0, 5)	SARIMA (0, 0, 2) x (2, 1, 3) [12]	ARIMA - ARCH (3)	ARIMA - GARCH (1,1)
Jan-2019	18.5	9.8	19.3	16.9	16.9
Feb-2019	33.1	3.9	23.3	17.4	17.5
Mar-2019	18.7	27.6	28.5	39.2	39.3
Apr-2019	31.5	77.1	38.9	78.6	78.6
May-2019	51.3	137.2	62.4	127.1	127.1
June-2019	113.5	186.9	170.2	167.5	167.4
July-2019	298.8	212.8	287.1	189.0	189.0
Aug-2019	299.9	207.9	255.5	186.2	186.2
Sep-2019	259.5	173.6	169.6	159.9	159.9
Oct-2019	110.1	119.1	74.3	117.1	117.1
Nov-2019	31.6	58.9	27.4	69.2	69.3
Dec-2019	19.2	9.3	14.7	28.9	29.0
Jan-2020	25.3	16.6	19.4	6.7	6.8
Feb-2020	10.8	11.7	21.5	8.4	8.4
Mar-2020	40.3	22.6	29.3	33.4	33.4
Apr-2020	46.2	77.2	38.5	75.0	75.0
May-2020	91.7	137.3	60.2	122.2	122.1
June-2020	196.2	187.0	162.6	162.5	162.4
July-2020	257.6	212.8	277.3	185.3	185.3
Aug-2020	327.0	207.9	250.4	184.8	184.7
Sep-2020	176.8	173.5	167.8	161.1	161.1
Oct-2020	78.1	119.0	69.9	120.6	120.6
Nov-2020	29.2	58.8	30.9	74.1	74.2
Dec-2020	17.0	9.2	13.6	33.8	33.9
Jan-2021	20.3	16.6	17.6	10.4	10.5
Feb-2021	7.6	11.6	22.8	9.9	9.9
Mar-2021	16.7	22.7	27.3	32.3	32.3
Apr-2021	31.0	77.3	37.3	71.6	71.6
May-2021	107.9	137.5	63.9	117.5	117.4
June-2021	182.9	187.1	167.7	157.6	157.5
July-2021	266.0	212.8	293.3	181.6	181.5
Aug-2021	196.2	207.9	256.4	183.2	183.1
Sep-2021	229.7	173.5	171.7	162.0	162.0

TABLE 2

The efficiency criterion of the traditional methods

Model	RMSE	MAE	Theils U	MAPE	Trend
ARIMA	8.1066	34.3979	0.2540	0..4844	0.7964
SARIMA	5.6094	22.5704	0. 1603	0.3054	0.9018
ARIMA-ARCH	9.3750	39.8743	0.3152	0.5476	0.7362
ARIMA-GARCH	9.3778	39.8852	0.3154	0.5478	0.7360

TABLE 3

The future predictions of the proposed models

	Observed	Simple Average Model	Weighted Average model	Linear Ensemble Model	Hybrid Stochastic model
Jan-2019	18.5	15.7	16.07	19.38	19.05
Feb-2019	33.1	15.5	16.29	23.45	22.69
Mar-2019	18.7	33.7	32.77	27.00	27.87
Apr-2019	31.5	68.3	64.39	34.10	37.88
May-2019	51.3	113.5	106.91	55.97	62.18
June-2019	113.5	173.0	172.98	167.95	167.96
July-2019	298.8	219.5	228.82	298.65	291.59
Aug-2019	299.9	208.9	215.47	262.28	257.16
Sep-2019	259.5	165.7	166.48	171.94	171.30
Oct-2019	110.1	106.9	102.62	71.29	75.28
Nov-2019	31.6	56.2	52.21	25.43	29.32
Dec-2019	19.2	20.5	19.33	13.82	14.93
Jan-2020	25.3	12.4	13.46	20.15	19.06
Feb-2020	10.8	12.5	13.75	23.93	22.70
Mar-2020	40.3	29.7	29.42	27.64	27.89
Apr-2020	46.2	66.4	62.76	34.36	37.89
May-2020	91.7	110.5	104.11	56.18	62.19
June-2020	196.2	168.6	168.27	167.62	167.91
July-2020	257.6	215.2	223.85	297.99	291.43
Aug-2020	327.0	207.0	213.09	261.89	257.08
Sep-2020	176.8	165.9	166.33	171.67	171.26
Oct-2020	78.1	107.5	102.52	70.56	75.24
Nov-2020	29.2	59.5	55.42	25.37	29.32
Dec-2020	17.0	22.6	20.99	13.30	14.92
Jan-2021	20.3	13.8	14.37	19.66	19.05
Feb-2021	7.6	13.6	14.81	23.93	22.70
Mar-2021	16.7	28.7	28.31	27.54	27.88
Apr-2021	31.0	64.5	60.97	34.53	37.89
May-2021	107.9	109.1	103.46	56.94	62.23
June-2021	182.9	167.5	168.04	168.54	168.04
July-2021	266.0	217.3	227.89	299.88	291.90
Aug-2021	196.2	207.6	214.50	262.62	257.23
Sep-2021	229.7	167.3	168.06	171.97	171.30

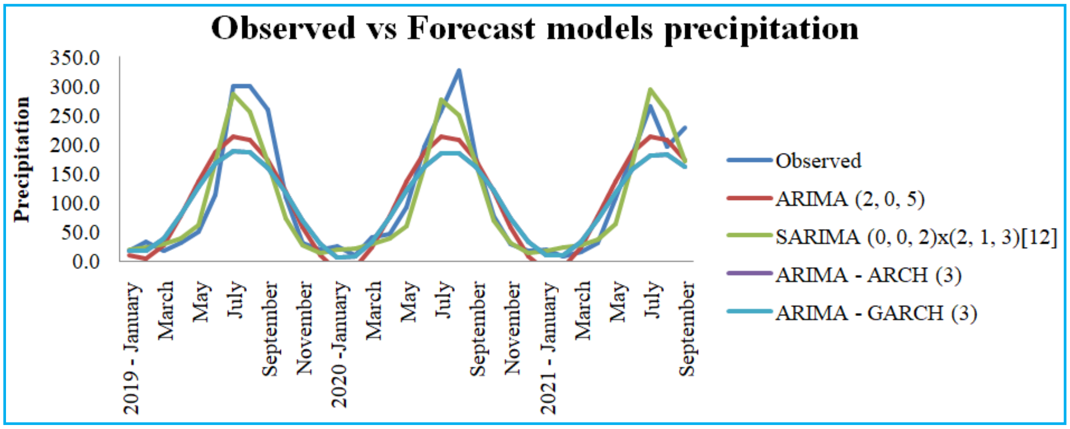

Fig. 1. Observed v/s Traditional forecast models precipitation

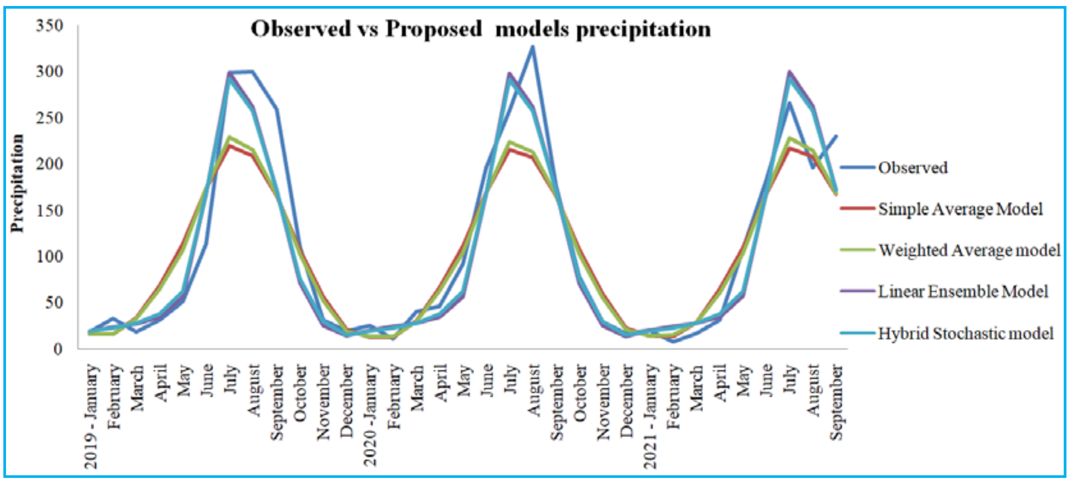

Fig. 2. Observed v/s proposed models precipitation

f_{ti} = the prediction of t^{th} model for i^{th} observation,

$\bar{f}_i$ = the mean prediction of the models for the i^{th} observation.

2.4. *Hybrid Stochastic Model* (*HSM*) : This model is complex than the other proposed models and in general, the higher the complexity, the greater the accuracy in the models. This model has constructed the inclusion of i^{th} cell probability for prediction.

$$\hat{y}_i = \overline{O}_i + p_i . \sum_{t=1}^{N} w_t . \left(f_{ti} - \bar{f}_i\right); t = 1,2...,N; i = 1,2,...m$$

where, p_i = the estimated probability for the i^{th} observation.

The weights are estimated based on the efficiency performance criterion of the forecasted models using normalization procedure.

Performance measures

Although the Root Mean Square Error (RMSE) is a good indication of performance. In reality, every testing technique's ultimate goal is to ensure that model findings are dependable and capable of evaluating model accuracy. To measure the prediction performance, the efficiency statistics Root Mean Square Error (RMSE), Theil's U

TABLE 4

The efficiency criterion of the proposed models

Model	RMSE	MAE	Theils U	MAPE	Trend
Simple Average Model	7.4963	30.7405	0.2387	0.4167	0.8311
Weighted Average Model	6.9961	28.6148	0.2198	0.3881	0.8535
Linear Weighted Average Model	5.6258	22.5315	0.1575	0.3118	0.9058
Hybrid Stochastic Model	5.4951	21.8998	0.1559	0.2968	0.9102

TABLE 5

Stepwise model selection

Model	RMSE	MAE	Theils U	MAPE	Trend
With one forecast (SARIMA) model	5.4846	21.8504	0.1561	0.2964	0.9109
With two forecasts (ARIMA & SARIMA models	5.4978	21.9168	0.1561	0.297	0.9103
With three forecast(ARIMA, SARIMA & GARCH) Models	5.4967	21.9082	0.1559	0.2969	0.9101
All models are included	5.4951	21.8998	0.1559	0.2968	0.9102

$(0 \le U \le 1)$, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Trend (R^2) component are utilized.

$$\text{RMSE} = \left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}\right]$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\left|y_i - \hat{y}_i\right|}{y_i}\right] * 100$$

$$\text{Theil's U} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}}{\left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^2\right)} + \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i\right)^2}\right]}$$

$$\text{Trend}\left(R^2\right) = \left[1 - \frac{V(e_t)}{V(y_t)}\right]$$

3. Data source and findings

For the analysis, we used 118 years of rainfall data to develop a viable time series model for predicting future rainfall. Monsoon winds have a huge impact on India's climate. The monsoon season, which lasts from June to September, brings the most rain to India. In this work, we collected data from the Indian Meteorological websites and journals to analyze 120 years of rainfall in India from 1901 to 2020. India has received an average of 97.85 mm of rainfall during the training years, with a low of 1.60 mm in December 1970 and a high of 375.5 mm in July 1959.

Findings

Initially, the traditional forecast models are used for prediction and tested with efficiency criteria to choose the desired model. Table 1 shows the predictions of these models. In general, to grasp the volatility of time series data, the ARCH/GARCH models will help. But on the other hand, the SARIMA showed better predictions than these models for seasonality and this is confirmed with the efficiency criterion from Table 2.

The precipitation is predicted for the test data with the traditional models and then the proposed models are constructed and also compared with the each other. Out of the four proposed models, Hybrid Stochastic Model showed a best prediction among the other models. The Table 3 shows the predictions of proposed statistical methods and the same is presented in Fig. 2.

The Hybrid Stochastic Model exhibited the best predictions and these results are confirmed with the comparative performance criteria presented in the Table 4. We also performed the model reduction technique to choose the best-forecasted models in yielding optimum results, *i.e.*, the selection of the forecasted techniques are utilized with the regression methods like All possible regression techniques, Backward and Forward stepwise approaches in the selection of the forecast models. All possible combinations of ($2^n = 16$; $n = 4$) of the forecasted models are verified with the proposed Hybrid Stochastic Model and the model with one forecast model (the SARIMA) resulted efficiently (this is confirmed with the performance measures in the Table 5, comparing to other models in terms of RMSE, MAE, Theil's U, MAPE and Trend component) and the best possible models in each step were presented in Table 5.

4. Conclusion

Initially, we pre-processed the data. The stationary of the data set is checked then we identified the appropriate model under the different scenarios. We made predictions with the traditional models. The SARIMA model showed very little biasedness in the estimation of precipitation for the test data. Further, we estimated the observed mean response or precipitation for each observation or timestamp, the weights of the models based on their performance criteria used in train data and the probability of each observation. Then, the future precipitation estimated by using the proposed statistical models; all the models are compared with the performance measures. The optimum number of forecasted models was identified with Hybrid Stochastic Model to estimate the best predictions using stepwise regression.

Declaration

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable at present

Availability of data and materials

All the data are available in the public domain.

Competing Interests

The authors declare that they have no competing interests.

Funding

Not applicable

Authors Contributions

Kumaraswamy developed the strategic model and drafted the paper. Bhatra Charyulu provided guidance and writing of the paper. Both authors contributed to the article in writing, reviewing, read and approving the submitted version of the manuscript.

Acknowledgments

We would like to thank the Editors as well as the unknown reviewers who have generously given up valuable time to review the manuscript with more insightful comments and suggestions which helped us to improve the quality of the original manuscript.

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Bates, J. M. and Granger, C. W. J., 1969, "The combining of forecasts", *Operational Research Quarterly*, **20**, 4, 451-468. doi : 10.1057/jors.1969.103.

Clemen, R., 1989, "Combining forecasts : A review and annotated bibliography", *International Journal of Forecasting*, **5**, 4, 559-583. doi : 10.1016/0169-2070(89)90012-5.

Kumaraswamy, K. and N. Ch. Bhatra Charyulu, 2019, "On Stochastic Distribution of Repeated Purchase Consumers", *Int. J. Agricult. Stat. Sci.*, **15**, 2, 841-846.

Kumaraswamy, K., 2022, "The Rainfall Forecast Models Analysis and Their Volatility", International, *Journal of Statistics and Reliability Engineering*, **9**, 3, 450-460. (ISSN(P) : 2350-0174; ISSN(O) : 2456-2378).

Kumaraswamy K. 2022, "Precipitation Trend Analysis of India- A Climate Change Study", *Indian Journal of Science and Technology*, **15**, 8, 351-356. https://doi.org/10.17485/IJST/v15i8.2040.

Kumaraswamy, K., Bhatracharyulu, N. Ch., 2023, "Statistical brand switching model : an Hidden Markov approach", *OPSEARCH* **60**, 942-950. https://doi.org/10.1007/s12597-022-00613-0.

Krishnamurti, T. N., Kishtawal, C. M., Zhan, Zhang, Timothy, Larrow, David, Bachiochi, Eric, Williford, Sulochana, Gadgil and Sajani, Surendran, 2000, "Multimodel Ensemble Forecasts For Weather And Seasonal Climate", *Journal of Climate*, **13**, 4196-4216.

Martins, V. L. M. and Werner, L., 2012, "Forecast combination in industrial series : A comparison between individual forecasts and its combinations with and without correlated errors", *Expert Systems with Applications*, **39**, 13, 11479-11486. doi : 10.1016/j.eswa.2012.04.007.

Poncela, P., Rodríguez, J., Mangas, R. S. and Senra, E., 2011, "Forecast combination through dimension reduction techniques", *International Journal of Forecasting*, **27**, 2, 224-237. doi : 10.1016/j.ijforecast.2010.01.012.