

Using Bayesian Model Averaging to Calibrate Temperature Forecast Ensembles in Hawassa Town, Ethiopia

WubshetBoru N*

Department of Mathematics, Arbaminch University, Arbaminch, Ethiopia

Abstract

Bayesian model averaging (BMA) is a statistical way of post-processing forecasts ensembles to create predictive probability density functions (pdfs) for weather quantities. BMA has been proposed as way of correcting under-dispersion in ensemble forecasts. The output of ensemble BMA is a weighted average of pdfs centered on the individual bias-corrected ensemble forecast. The BMA weights are posterior probabilities of the models generating the forecast, and show the relative contribution of the component models to the prediction over the training period. This paper focuses on short range forecasting of daily maximum temperature. Forecasting temperature, we have approximated the conditional pdf by normal distribution centered at a linear function of the forecast. Since we are not able to get an ensemble of forecasts for Ethiopia, we have used the previous eight years from 2003 to 2010 data as ensemble and 2011 as observation. The BMA approach yields predictive distribution for temperature that were much better calibrated than those based on naive averaging of the forecasts.

Keywords: Bayesian model averaging; Meteorology; Forecasting; Ensemble; Forecast uncertainty

Introduction

History tells that people have been systematically recording and comparing meteorological data for some 250 years. But as time goes and development of different technologies quickly changes, it has become very important and essential to provide meteorological forecasts for the near future [1].

The weather on Earth is governed by scientific laws admittedly, very complex and imperfectly understood laws. That is why it is very difficult to forecast the weather for any one place, or over a very long period, with the type of accuracy we associate with the exact sciences. Thus forecast must be based on rational study of the phenomenon, space and time series data of weather parameters and analyses of correlated meteorological conditions.

During the last decade, multi-model ensemble prediction systems have become the basis for probabilistic weather and climate forecasts at many operational centers throughout the world. Multi-model ensemble predictions aim to capture several sources of uncertainty in numerical weather forecasts, including uncertainty about the initial conditions, lateral boundary conditions and model physics, and have convincingly demonstrated improvements to numerical weather and climate forecasts and production of more skillful estimates of forecast pdf. However, because the current generation of ensemble systems do not explicitly account for all sources of forecast uncertainty, some form of postprocessing is necessary to provide predictive ensemble pdfs that are meaningful, and can be used to provide accurate forecasts.

Bayesian model was originally developed as a way to combine inferences and predictions from multiple statistical models, and was applied to statistical linear regression and related models in applied social and health sciences [2].

Raftery et al. have extended BMA to ensembles of dynamical models and showed how it can be used as a statistical post processing method for forecast ensembles, yielding calibrated and sharp predictive pdfs of future weather quantities [3]. The BMA predictive pdf of any future weather quantity of interest is a weighted average of pdfs centered on the individual ensemble members. They have applied the method to

surface temperature and sea level pressure and the result of their study showed that the predictive pdf were much better calibrated than the raw ensemble and BMA yields a deterministic point forecast.

Bayesian model averaging [4] have been used by many scholars in different countries and yields consistent improvement in forecast performance.

This study focus on short term weather forecasting, which can be one day or one week ahead by using BMA to make forecast and to assess forecast uncertainty in temperature forecasting at Hawassa City SNNPR, Ethiopia.

Data and Methodology

This study has been conducted in Hawassa is a city, on the shores of Lake Hawassa in the Great Rift Valley. Located in the Sidama Zone 270 km south of Addis Ababa via DebreZeit, 130 km east of Sodo, 75 km north of Dilla and 1125 km north of Nairobi, Hawassa is the capital of the Southern Nations, Nationalities, and Peoples Regional State. The city lies on the Trans-African Highway 4 Cairo-Cape Town, with a latitude and longitude of 73° N 3828' E Coordinates: 73° N 3828' E and an elevation of 1708 meters.

Data description

In order to apply the BMA method we used a secondary data from the National Meteorological Agency (NMA) in Ethiopia. This includes real daily observations from Hawassa station for nine years. However, regarding the ensemble data, the NMA is not currently using multi-models and ensemble data as we stated in the introduction, analogue

*Corresponding author: WubshetBoru N, Professor, Department of Mathematics, Arbaminch University, Arbaminch, Ethiopia, Tel: +251 46 881 4986; E-mail: nitsuh.wubshet@amu.edu.et

Received Mar 13, 2018; Accepted April 30, 2018; Published May 02, 2018

Citation: WubshetBoru N (2018) Using Bayesian Model Averaging to Calibrate Temperature Forecast Ensembles in Hawassa Town, Ethiopia. J Phys Math 9: 267. doi: 10.4172/2090-0902.1000267

Copyright: © 2018 WubshetBoru N. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

methods are used which mean that it is impossible to obtain the ensemble data from Ethiopia.

We have chosen to collect historical data from NMA for several years and will later explain how we have adapted this to the ensemble setting needed to perform BMA. The historical data obtained from NMA consists of daily observations of maximum temperature Hawassa city for 2003-2011 [5-8]. The data is presented according to the Gregorian calendar starting from January 1st to December 31st and discussed in the rest of this section.

Maximum temperature

The temperature data obtained from NMA, expose yearly seasonal variations – higher in June-August and lower in December-February. Otherwise, there seems to be fairly erratic variations (unpredictable), possibly with some short-term correlations.

Methodology

Consider y the quantity of interest to be forecasted on the basis of training data (set of data in which we observe the outcome and feature measurements for a set of objects) and by using K statistical models M_1, \dots, M_k .

Consider $[y_1, y_2, \dots, y_T, Y_{T+1}]$ where the first T terms are observed, let it be denoted by y^0 and the other term at $T+1$ is the forecast, let it be denoted by y . In other words the forecasting variable can be written as $[y^0, y]$.

Also consider forecast ensemble from the K models $[f_{k1}, \dots, f_{kT}, f_{kT+1}]$ where $k=1, \dots, K$. The reference t represents time interval between each observation, which is normally one day or one week, and which is the same for y and f variables.

Bayesian model averaging

The traditional idea is that at any particular time, there is one best ensemble member, and that if we knew which it was, we would use that member as forecast. Hence we will not take account of that the uncertainty in identifying that ensemble, and this is accounted for through Bayesian model averaging. BMA can be used to calibrate forecast ensembles and is currently being implemented in many weather centers. The BMA can be defined either with a model or an ensemble focus [9-15].

Model

Consider K forecast models, M_1, \dots, M_k . For the quantity y to be forecasted on the basis of data y^0 using K statistical models, the law of total probability tells us the forecast pdf, $P(y)$ named BMA pdf. The best forecast model, is given by

$$P(y) = \sum_{k=1}^K P(y | M_k) P(M_k | y^0) \quad (1)$$

Where, $P(y|M_k)$; $k=1, \dots, K$ is forecast pdfs based on each model M_k alone, and $P(M_k|y^0)$ is posterior probability of model M_k being correct given the training data y^0 . It reflects how well model M_k fits the training data. The BMA pdf is a weighted average of the conditional pdfs given each of individual model, weighted by their posterior model probabilities.

Ensemble focus

Ensemble generated values from slightly different models of the climate system. Ensemble forecasting is a numerical prediction method that is used to attempt to generate a representative sample of

the possible future states of a dynamical system. Monte Carlo analysis is a means of statistical evaluation of mathematical functions using random samples. Ensemble forecasting is a form of Monte Carlo analysis: multiple numerical predictions are conducted using slightly different initial conditions that are all plausible given the past and current set of observations, or measurements. Sometimes the ensemble of forecasts may use different forecast models for different members, or different formulations of a forecast model. The multiple simulations are conducted to account for the two sources of uncertainty in weather forecast models: (1) The errors introduced by chaos or sensitive dependence on the initial conditions; and (2) errors introduced because of imperfections in the model. Ideally, the verified weather pattern should fall within past ensemble spreads, and the amount of spread should be related to the probability of certain weather events occurring.

The basic idea here is that for any given forecast ensemble there is a best model, or member, but we do not know what it is, and our uncertainty about the best member is quantified by BMA. Each deterministic forecast can be bias-corrected using any one of many possible bias corrected methods, yielding a bias-corrected forecast. Then the f_{kt} is linked to y_t by a conditional pdf of $[y_t|f_{kt}]$. This is the conditional pdf of y given that f_k is the forecast in the ensemble. Consider K forecast ensembles: $[f_{k1}, \dots, f_{kT}, f_{kT+1}]$ where $k=1, \dots, K$. Hence the bias-corrected forecast ensembles are;

$$y_t = a_k + b_k f_{kt} + E \quad (2)$$

For $t=1, \dots, T, T+1$ and $k=1, \dots, K$, where $E \sim N(0, \sigma^2)$

When forecasting temperature, it is often reasonable to approximate the conditional pdf by a Gaussian distribution centered at a linear function of the forecast, hence

$$[y_t | f_{kt}] \sim N(a_k + b_k f_{kt}, \sigma^2) \quad (3)$$

Hence forecast y based on f_k for $k=1, \dots, K$ can be written as follows.

$$[y | f_1, \dots, f_k] = \sum_{k=1}^K w_k [y | f_k] \sim N\left(\sum_{k=1}^K w_k [a_k + b_k f_k], \sigma^2 \sum_{k=1}^K w_k^2\right) \quad (4)$$

Where; w_k, a_k, b_k for $k=1, \dots, K$; and σ^2 are model parameters. The w_k is the posterior probability of the forecast k being the best one and is based on the forecast performance in the training period; see also expression (1).

For temperature forecasting, since the normal distribution is appropriate, we can apply directly the above method [3]. The best forecast for y (in our first case y is temperature) is then:

$$\mu_{y|f} = E\{y | f_1, \dots, f_k\} = \sum_{k=1}^K w_k (a_k + b_k f_k) \quad (5)$$

This can be viewed as a deterministic forecast in its own right and can be compared with the individual forecast ensembles, or with the ensemble mean. The predictive variance of y can be written as:

$$\sigma_{y|f}^2 = Var\{y | f_1, \dots, f_k\} = \sum_{k=1}^K w_k ((a_k + b_k f_k) - \mu_{y|f})^2 + \sigma^2 \quad (6)$$

This variance can be seen as:

Predictive Variance = Between-Forecast Variance + Within-Forecast Variance

Hence as we can see that the right hand side has two terms, the first part summarizes between-forecast (ensemble) spread. The ensemble spread alone may underestimate uncertainty, since it ignores the second term on right-hand side which is the expected uncertainty conditional on one of the forecast being best. Therefore BMA capture the spread-

error variable, but also accounts for the possibility that ensembles may be under dispersive.

BMA provides a theoretical framework for understanding these apparently contradictory effects and suggests ways to correct for them. The parameters of the model, namely $(w_k; a_k; b_k$ for $k=1, \dots, K$, and σ^2) are unknown and must be estimated.

Parameter estimation

Here we consider estimate of the model parameters based on the training dataset consisting of the ensemble forecast and the verifying observations. If the forecasts have not yet been bias corrected, $a_k; b_k$ for $k=1, \dots, K$, can be estimated by simple linear regression of $y^0, f_k^0; k=1, \dots, K$ for the training data.

Moreover σ^2 and w^k for $k=1, \dots, K$ can be estimated by maximum likelihood from y^0, f_k^0 where $k=1, \dots, K$ from the training data. The likelihood model is defined as the probability of the training data given the parameter to be estimated. The maximum likelihood estimator is the value of the parameter vector that maximizes the likelihood function, that is, that value of parameter vector under which the observed data where most likely to have been observed. The log-likelihood function to be used in optimization is:

$$l(\sigma^2, w_k; y^0, f_k^0) = \sum_{t=1}^T \log \sum_{k=1}^K w_k \phi(y_t^0; \hat{a}_k + \hat{b}_k f_{kt}^0, \sigma^2) \quad (7)$$

for $k=1, \dots, K$. Where $\phi(\cdot; \mu, \sigma^2)$ is the Gaussian pdf, y^0 is the observed values .

The maximum likelihood estimator is defined as:

$$[\hat{\sigma}^2, \hat{w}_k] = \underset{\sigma^2, w_k}{\operatorname{argmax}} \{l(\sigma^2, w_k; y^0, f_k^0)\} \quad (8)$$

Where, f^0 is the ensemble forecasts up to t . This cannot be maximized analytically, but we maximize it numerically using the expectation-maximization (EM) algorithm.

EM algorithm

The EM algorithm is method for finding the maximum likelihood estimator when the problem can be recast in terms of unobserved quantities such that, if we knew what they were, estimation would be straightforward.

The BMA model in expression 4 is a finite mixture model. We introduce the unobserved quantities z_{kt} , where $z_{kt}=1$ if ensemble member k is the best forecast for verification time t , and $z_{kt}=0$ otherwise. For each t , only one of z_{1t}, \dots, z_{Kt} is equal to 1; the others are all zero. The EM algorithm is iterative and alternates between two steps, the E (or expectation) step and the M (or maximization) step. It starts with an initial guess, Θ^0 , for the parameter vector.

In the E step, the z_{kt} are estimated given the current guess for the parameters; the estimates of the z_{kt} are not necessarily integers, even though the true values are 0 or 1. In the M step, Θ is estimated given the current values of the z_{kt} . For the normal BMA model given by expression (3) and (4), the E step is;

$$\hat{z}_{kt}^{(j)} = \frac{w_k \phi(y_t | f_{kt}, \sigma^{(j-1)})}{\sum_{i=1}^K w_i \phi(y_t | f_{it}, \sigma^{(j-1)})} \quad (9)$$

Where the superscript j refers to the j^{th} iteration of the EM algorithm, and $\phi(y_t | f_{kt}, \sigma^{(j-1)})$ is normal pdf with mean $\hat{a}_k + \hat{b}_k + \hat{b}_k f_{kt}$ and standard deviation $\sigma^{(j-1)}$ evaluated at y_t . The M step then consists of estimating the w_k and σ using as weights the current estimates of $z_{kt}^{(j)}$, namely $\hat{z}_{kt}^{(j)}$. Hence,

$$w^{(j)} = \frac{1}{T} \sum_t \hat{z}_{kt}^{(j)}, \quad (10)$$

$$\sigma^{2(j)} = \frac{1}{Tk} \sum_t \sum_{k=1}^K \hat{z}_{kt}^{(j)} (y_t - \hat{a}_k - \hat{b}_k f_{kt})^2 \quad (11)$$

Where; T is the number of observations in the training set.

The E and M step are then iterated to convergence, which we define as no changes not greater than small tolerance in any of the log likelihood, the parameter value or the $\hat{z}_{kt}^{(j)}$ in one iteration. The log likelihood is expected to increase at each EM iteration which implies that in general it converges to a local maximum of the likelihood. Convergence to a global maximum cannot be guaranteed, so the solution reached by the algorithm can be sensitive to the starting values. Starting values based on past experience usually give good solutions [16-19].

Assessing calibration

Calibration refers to statistical consistency between the predictive distribution and observation [4]. The ensemble function for verification can be used whenever observed weather condition is available. This function computes the mean absolute error, continuous ranked probability score. The mean absolute error (MAE) computes the mean absolute difference between the ensemble or BMA median forecast and the observation, whereas the CRPS is the integral of Brier scores.

Results and Discussion

The temperature data is analyzed as follows. For each season Predict temperature of last day in each season of 2011-that is y .

Two types of training data; y_0 :

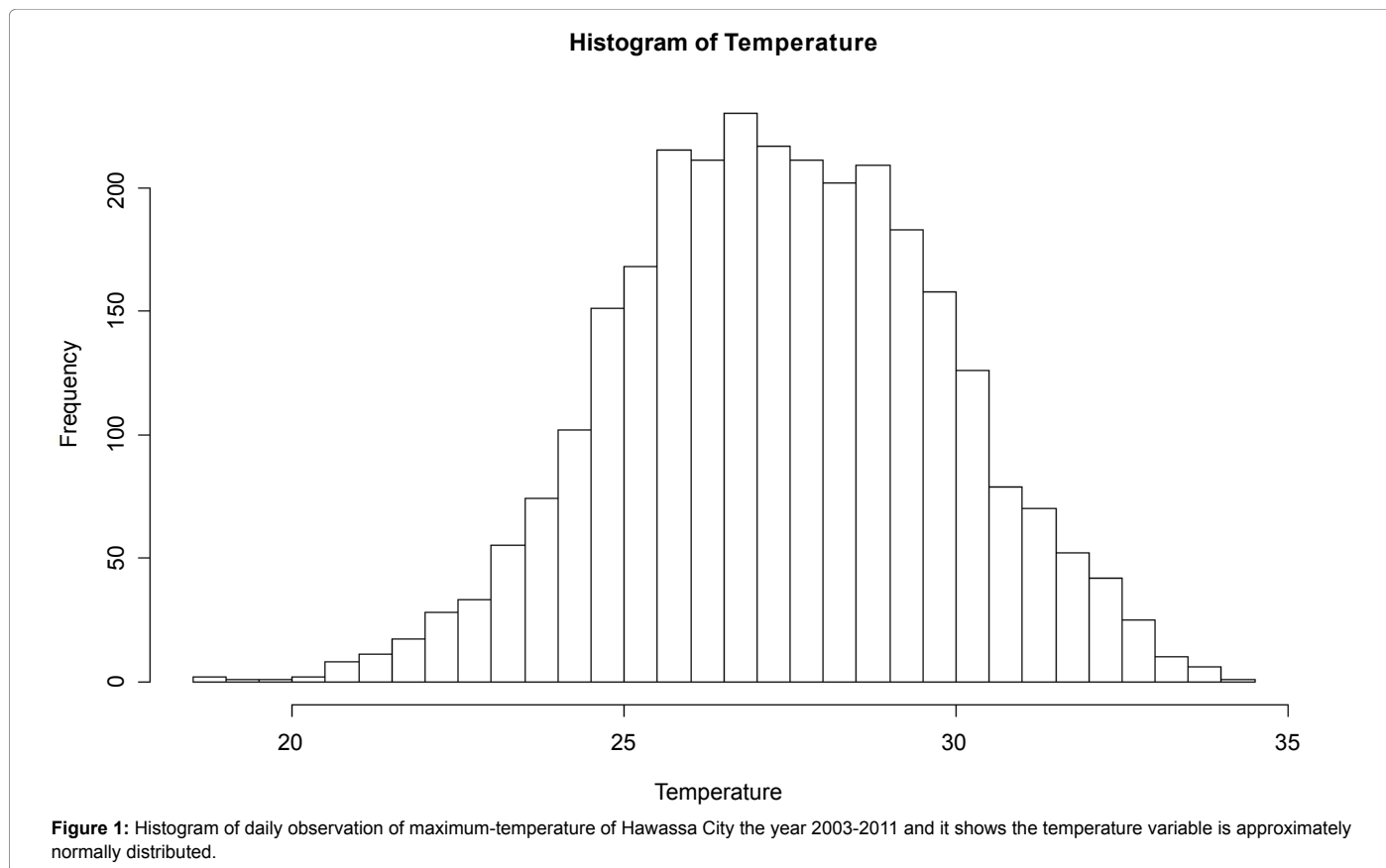
1. Three first months of season- average of month, and average of the first fifteen days of last month. Daily observations from last half of last month. This provides 4+15=19 training data for each season in 2011.
2. Average of the first fifteen days of month and daily observation from last half of last month of the season. This provides 1+15=16 training data for each season.

Ensembles are defined to be the temperature profile for the same season through the years 2003- 2010. This provides 8 ensemble members f_1, \dots, f_8 . Note that NMA do not provide forecasts ensembles, hence we have chosen to use historical data for the same season as the ensemble. This approach has been recommended in Hamill et al. [15].

We will provide Ensemble BMA temperature forecast for three days-the last day of each season. We will provide three forecasts for each day. Based on training data, set 1 and 2 respectively, and a naive average of the same day from the eight ensemble members, Prediction variance is also estimated.

We start with data set-1 for each season and continue data set-2. The results are presented in Figures 1, 2 and Table 1. But for this paper we only present the result for the January 31st using both data set 1 and 2, respectively. Finally, the actual temperature prediction with prediction variance is presented. First we present the results for January 31st with training data set 1, in Figure 2A.

Figure 2A the individual eight ensemble forecast and observed values for Bega (January 31st) are displayed. From this figure, we can see that the observed values (shown by dots) are almost inside the ensemble. However, there is variability in closeness of the ensemble to the corresponding true value. To correct, then we did the bias correction



(i)		ens1	ens2	ens3	ens4	ens5	ens6	ens7	ens8
	a_k	21.725	27.136	40.3	16.97	43.474	23.606	24.266	6.042
	b_k	0.26	0.073	-0.388	0.399	-0.493	0.191	0.174	0.793
(ii)		ens1	ens2	ens3	ens4	ens5	ens6	ens7	ens8
	w_k	6.61e-21	2.85e-25	3.33e-01	1.80e-06	1.18e-06	1.01e-22	6.75e-28	4.87e-01
	sig	0.485							
(iii)		ens1	ens2	ens3	ens4	ens5	ens6	ens7	ens8
	a_k	17.692	28.735	41.312	5.485	44.02	23.989	24.087	6.546
	b_k	0.395	0.013	-0.423	0.446	-0.515	0.176	0.177	0.772
(iv)		ens1	ens2	ens3	ens4	ens5	ens6	ens7	ens8
	w_k	4.56e-15	3.35e-56	2.03e-01	2.40e-02	2.30e-49	5.20e-49	4.48e-55	5.35e-01
	sig	0.504							
(v)	Data	True value	BMA forecast	Pre. variance	naïve forecast		var. of naïve		
	set-1	31	30.077	30.077	0.485		29.187		
	set-2	31	30.026	30.026	0.503		29.216		

Table 1: (i) Bias-coefficients for January 31st of data set 1; (ii) Weights given to ensembles for January 31st of data set 1; (iii) Bias-coefficients for January 31st of data set 2; (iv) weights given to ensembles for January 31st of data set 2; (v) BMA for January 31st.

and it is displayed in Figure 2B. The bias correction coefficients are listed in Table 1(i). Here negative sign indicates ensemble is negatively corrected with the training data. This may happen when real historical ensembles are used. If we assume that the pattern is equal for increasing and decreasing temperature profile, this negative correlation is acceptable.

Furthermore, after the bias correction, weighted bias corrected forecast were calculated. Hence this is shown in Figure 2C by thickness of the line. The more thick the line the more it is closer to the true values. The weights are given to each ensemble according to their power to forecast the truth values.

We have ensemble forecasts day $t+1$ and observed values for t days. Our aim is to forecast the last values $t+1$. Each figure contains several displays. The first display presents the training y^0 and the eight ensembles f_1, \dots, f_8 . The next display is the bias corrected ensemble members $y_i, i=1, \dots, 8$. The actual coefficients ($a_k; b_k$); $k=1, \dots, 8$ are also listed. The last display present the bias-corrected ensembles with line-thickness proportional to the weights ($y_i; w_k$); $i=1, \dots, 8$. Lastly, the actual weights are listed with the associated estimated variance. We have used the ensemble BMA package in the package in R library.

Hence more weights is given to ensemble 8 and 3, in our case data from 2010 and 2004 respectively. As consequence, from our results we

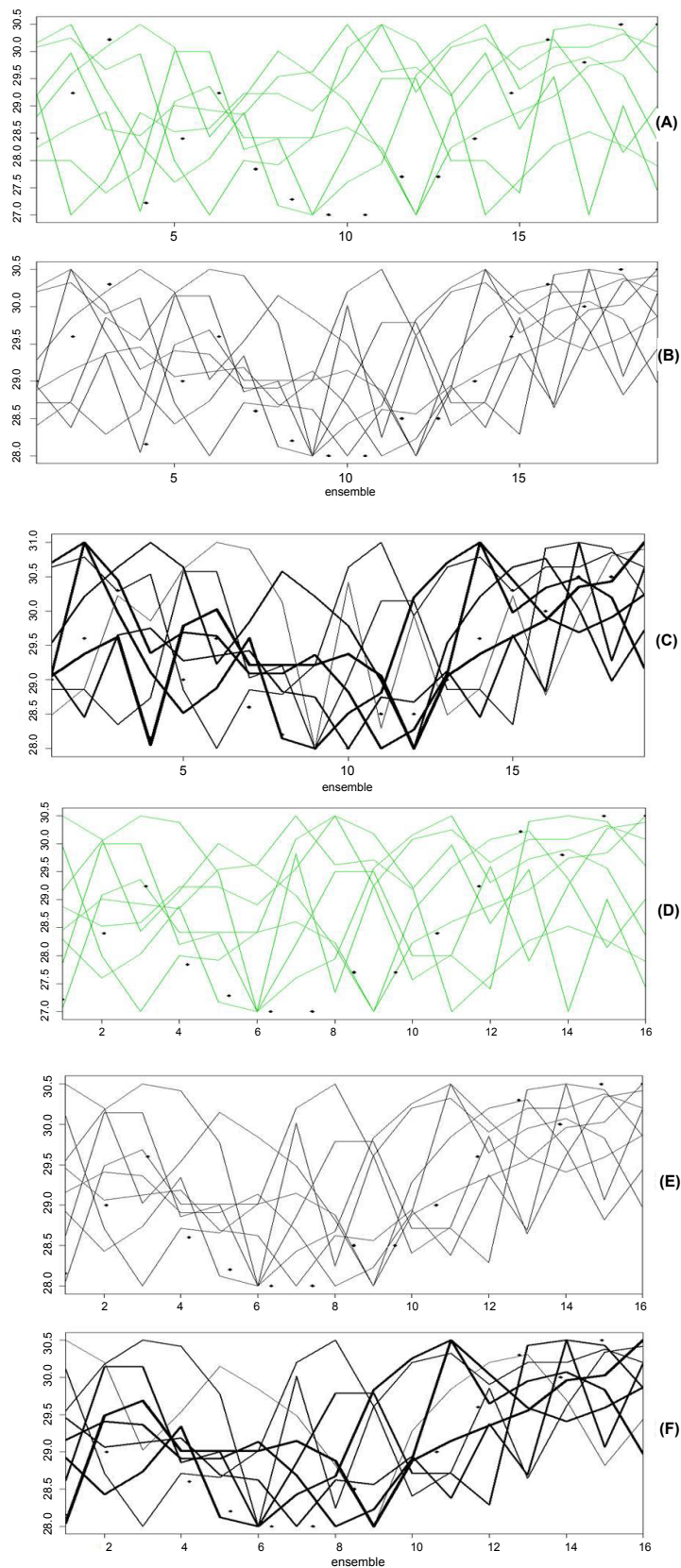


Figure 2: (A) the eight ensembles and the training data set 1 for January 31st; (B) The Bias-corrected ensembles for January 31st; (C) Weighted-Bias-corrected ensembles for January 31st; (D) The eight ensembles and training data set 2 for January 31st; (E) Bias-corrected ensembles for January 31st; (F) Weighted Bias-corrected ensembles for January 31st.

may say that the season in these years is more closed to the weather condition of the corresponding season of the year 2011. But we should know every ensemble has its own weights even if their power may be small. The estimated weights are listed in Table 1(iii). Some weights are very close to zero, which shows that the ensemble have less predictive power. Note that the sum of the weights is always equal to one.

The next Figures 2D-2F displays the results from the analysis based on training data set 2, and they display almost similar behavior to the corresponding result of dataset 1. Our main reason to use data set 2 is to see changes in the value of the forecast with different length of the training data. However, there was no dramatic change in the values of bias coefficients and weights values.

Table 1(v) show the true values, BMA forecast and the corresponding prediction variance, and finally, the naive forecast and its variance. The BMA forecast look reliable than the corresponding naive ones closer to true value in almost cases and considerably less prediction variance which still is representative. The true value is well within two standard deviations of the BMA forecast. Forecast based on data set 2, the one month data set, does as well as the four month data set, hence the one month training has smaller variance and better to use ensemble forecast.

Conclusion

In this study we have applied a normal ensemble BMA to forecast the temperature at the last day of the three seasons, Bega, Belg and Kiremt. Two different conditioning schemes are used. One of which is four month long and one month long. From our study it seems that forecasts based on four months of observations are only slightly more reliable than the one-month forecast. These forecasts are also superior to naive forecasts based on the average of previous years temperature. The forecast variance appears reliable. Hence from our study, we can conclude that Bayesian model averaging yields much better and calibrated results than the individual forecast ensembles and the naive forecasts.

From our study we have argued that it is important to take account of model uncertainty when making forecast. A coherent and conceptually simple way to do this is Bayesian model averaging; it provides better average predictive performance than any single model that could be selected. Bayesian model averaging also avoids the problem of having to defend the choice of any particular model, thus simplifying the presentation of the results.

References

1. Weather and Climate Forecasts for Agriculture.

2. Hoeting J, Madigan D, Raftery A, Volinsky C (1999) Bayesian Model Averaging A Tutorial. *Journal of Statistical Science* 14: 382-417.
3. Raftery A, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Journal of American Meteorological society* 133: 1155-1174.
4. Sloghter J, Raftery A, Gneiting T, Fraley C (2007) Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Journal of American Meteorological Society* 135: 3209-3220.
5. Barnston A, Diriba K (2007) Predictability of June-September Rainfall in Ethiopia. *Journal of American Meteorological Society* 135: 628-650.
6. Berrocal V, Raftery A, Gneiting T (2006) Combining Spatial and Ensemble information in probabilistic weather fore-casts. *American Meteorological Society*.
7. Buizza R, Houtekamer P, Toth Z, Gerald P (2005) A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Journal of American Meteorological Society*.
8. Cheung W, Gabriel B, Senay G, Singh A (2008) Trends and spatial distribution of annual and seasonal rainfall in Ethiopia. *National Journal of Climatology* 28: 1723-1734.
9. Fraley C, Raftery A, Gneiting T, Sloghter J, Berrocal V (2011) Probabilistic Weather Forecasting in R. *The R Journal* 3: 55-63
10. Gneiting T, Balabdaoui F, Raftery A (2005) Probabilistic Forecasts, Calibration and Sharpness. *Royal Statistical Society*.
11. Jones J, Hansen J, Royce F, Messina D (2000) Potential benefits of climate forecasting to agriculture. *Journal of Agriculture, Ecosystems and Environment* 82: 169-184.
12. Hutter K (2011) *Advances in Geophysical and Environmental Mechanics and Mathematics*. Springer.
13. Steinbakk G (2008) Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology* 44: 92-107.
14. Tapiador F, Turk F, Petersen W, Hou A, Garca-Ortega E, et al. (2012) Global Precipitation Measurement: Methods, Datasets and Applications. *Journal of Atmospheric Research* 104-105: 70-97.
15. Hamill TM (2004) Ensemble Based Atmospheric data Assimilation. Technical Report 1-47.
16. Gneiting T (2008) Bayesian Model Averaging and the Geo-statistical Output Perturbation technique: Statistical approaches to probabilistic weather forecasting. University of Washington.
17. Unger D, Dool H, O Lenic E, Collins D (2009) Ensemble Regression. *Journal of American Meteorological Society* 137: 2365-2379.
18. Vrugt J, Robinson B (2007) Treatment of Uncertainty using Ensemble Methods: Comparison of sequential data assimilation and Bayesian model averaging. *Journal of the American Geophysical Union* 43: 1-15.
19. Wilson L, Beauregard S, Raftery A, Verret R (2007) Calibrated Surface Temperature Forecasts from the Canadian Ensemble Prediction System Using Bayesian Model Averaging. *Journal of American Meteorological Society* 135: 1364-1385.