MATH11188 Statistical Research Skills: Assignment 1
Teng Wei Yeo, S2566430

1 Introduction

This report highlights the novel scientific contributions of Sloughter *et al.* (2010), explores its limitations, and creates a small experiment with simulated discretised data from a gamma distribution to compare the alternative BMA methods.

2 Key contributions

2.1 Overview

This paper uses Bayesian Model Averaging (BMA) to derive prediction probability density functions (PDFs) of maximum wind speeds from forecasts made by the University of Washington mesoscale ensemble.

Prediction PDFs are more informative than point predictions. To approximate prediction PDFs, previous literature have used (i) the range of predictions made by an ensemble as a measure of uncertainty, or (ii) post-processing using quantile regressions and computing prediction intervals at different quantiles. Method (i) is unsuccessful because observations sometimes fall outside the range of the ensemble predictions, and (ii) is computationally inefficient.

BMA, developed by Raftery *et al.* (2005), directly calculates a combined predictive PDF conditional on the forecasts of each ensemble member $k = 1, ..., K$. The BMA PDF is a weighted average of the individual forecasts,

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

where $p(y|f_1, ..., f_K)$ is the BMA PDF, and $g_k(y|f_k)$ are the prediction PDFs of each member. Weights w_k are adjusted based on the success of each member's past predictions relative to observed data.

2.2 Methodological Novelty

2.2.1 Discretisation in BMA

Estimation of the hyperparameters of the individual predictive PDFs, and the weights w_k, can be done by maximum (log) likelihood estimation using past data,

$$\ell(w_1, ..., w_K; c_0, c_1) = \sum_{s,t} \log p(y_{st}|f_{1st}, ..., f_{Kst}), \tag{1}$$

where s, t are location and time.

The key insight in Sloughter et al. (2010) is in addressing the fact that wind speeds y_{st} can be zero. Calculating (1) becomes untenable when y_{st} is zero if the log likelihood contains $\log(y_{st})$. They hence apply the following substitution (as developed by Wilks (1990)),

$$p(y_{st}|f_{1st}, ..., f_{Kst}) = \int_0^1 p(y|f_{1st}, ..., f_{Kst}) \, \mathrm{d}y, \tag{2}$$

when y_{st} is zero. The authors refer to this as the *Standard Method*. The model hence directly accounts for how observed wind speeds below 1 knot are rounded down to zero.

The other novel contribution is the generalisation to a *Fully Discretised Method*, where the substitution done in (2) is repeated for *all* the observed integer wind speed values y_{st}, and the limits of integration are the values which round to that integer. For example, $p(y_{st} = 3|f_{1st}, ..., f_{Kst}) = \int_{2.5}^{3.5} p(y|f_{1st}, ..., f_{Kst}) \, \mathrm{d}y$. They also develop the *Doubly Discretised Method*, where the forecasts are also discretised.

2.2.2 Fast MLE Computation

The MLE for each discretised BMA is computed using the ECME algorithm (Liu and Rubin 1994). At iteration $(j+1)$, the E step finds $\hat{z}_{kst}^{(j+1)}$ (i.e. the probability that f_k is the best forecast for the observed value at (s, t)) given the current iteration's weights and hyperparameters:

$$\hat{z}_{kst}^{(j)} = \frac{w_k^{(j)} p^{(j)}(y_{st}|f_{kst})}{\sum_{l=1}^{K} w_l^{(j)} p^{(j)}(y_{st}|f_{kst})}. \tag{3}$$

The CM-1 step then computes

$$w_k^{(j+1)} = \frac{1}{n} \sum_{s,t} \hat{z}_{kst}^{(j)}, \tag{4}$$

which intuitively adjusts the weights on each member's PDF using the average of the probability that the member performed the best at each observation at (s, t). The CM-2 step then conditions on these updated weights to find the MLE for the other hyperparameters.

A significant result is that the slow CM-2 step can be done infrequently (once every fifty E and CM-1 iterations) yet still reach similar results as to when the CM-2 step is repeated every iteration. Future research using ECME can consider a similar approach to reduce computation time.

2.3 Application-specific Novelty

2.3.1 Goodness of Fit of Gamma Model

Sloughter et al. (2010) show that a gamma distribution is an appropriate choice for the functional form of the conditional BMA PDF $p(y|f_1, ..., f_K)$ model for maximum wind speeds.

Figure 1 shows the exceptional fit of the gamma distribution to the non-transformed observations (Sloughter *et al.* (2010), p. 28, Figure 2(b)).

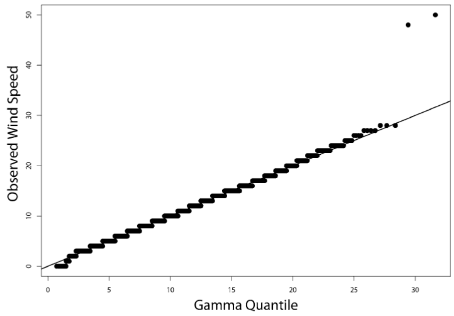

Figure 1: From Sloughter *et al.* (2010), p. 28, Figure 2(b). Gamma quantile-quantile plot of actual maximum wind speeds when member 1 forecasts 5-10 knots.

Their analysis also showed that the mean and standard deviation (s.d.) of the gamma distribution fitted to observed values are linearly dependent on the forecasted values. This hence justifies linearising the relationship between forecasts, and the mean and s.d. of the gamma distributions:

$$\mu_k = b_{0k} + b_{1k} f_k$$
$$\sigma_k = c_{0k} + c_{1k} f_k$$

2.3.2 Success of the Parsimonious Model

The BMA methods perform better than climatology and ensemble forecasts of maximum wind speeds. BMA methods have smaller forecasting error and better calibration (i.e. a 77.8% prediction interval more likely includes observed values) than ensembles, and is sharper (prediction interval is smaller) than climatology.

Significantly, they showed that all the methods of BMA used, as described in §2.2.1, have similar performance in calibration and sharpness. Even a parsimonious model, which assumes that all forecast models have the same mean and s.d. parameters b_0, b_1, c_0, c_1, still performs very well.

3 Limitations

3.1 Lack of Generalisability

Many of this paper's key contributions are specific to wind-speeds. The insight that the parsimonious model performs well may not hold in other applications, as noted by the authors.

Future research should consider confirming that the goodness of fit of the gamma model, the linear relationships of the mean and s.d. to the forecasts, and the success of the parsimonious model remain valid for wind speeds at different locations.

3.2 Climatology Might Be Superior

The BMA likely performs worse than climatology if there is substantial geographic changes at specific stations which are not captured by the grid resolution of the ensemble. Sloughter *et al.* (2010) use the University of Washington mesoscale ensemble, a 12km grid over the Pacific Northwest (Eckel and Mass, 2005). Gneiting *et al.* (2008) show that estimation from grid-based forecasts (like BMA) perform worse than climatology on forecasts at West Vancouver, British Columbia, because this station is close to the ocean, and has local land-water effects not captured in the 12km grid averages.

4 Extensions

4.1 BMA with Climatology

The implication of §3.2 is to instead use an aggregated BMA-with-Climatology approach. This can perhaps be done by fitting a gamma distribution over the climatology forecasts $g_{k+1}(y|f_{k+1})$ where $k+1$ is the climatology member, then computing $p(y|f_1, ..., f_K, f_{K+1}) = \sum_{k=1}^{K+1} w_k g_k(y|f_k)$, deriving the weights using the method in §2.2.2.

4.2 Higher-resolution Forecasts

A limiting factor of BMA is the resolution of the forecasts. This can perhaps be improved by making a continuous prediction distribution over the whole space. Possibilities include using a Generalised Additive Model with a thin plate spline (Wood and Augustin, 2002), or using R-INLA to compute a Gaussian Field over the continuous space (Lindgren and Rue, 2015), although the latter, which assumes normality, would not be appropriate for wind speeds.

5 Toy Experiment

I ran a simulation to test the difference between the *Standard Method* and the *Fully Discretised Method*. I assume only 1 location and 4 ensemble members. Prediction PDFs of each member and the hyperparameters are assumed to be known. Using the E and CM-1 steps, I derive the weights for the BMA PDF trained on 20 observations of training data. The goal is to compare the shape of the BMA PDFs from the *Standard Method* versus the *Fully Discretised Method*.

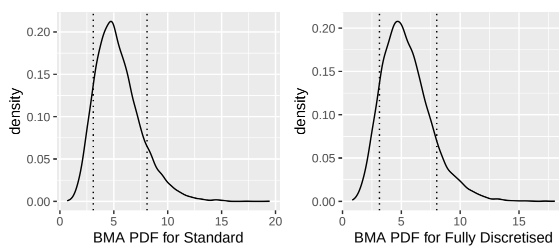

Figure 2: BMA PDFs with 77.8% Pred. Interval

Figure 2 shows that the BMA PDFs for each method, as derived in my simulation, are almost identical, confirming the results in Sloughter et al. (2010). My code is at: `https://github.com/affort/SRS_A1_S2566430.git`.

References

Eckel, F. A. and Mass, C. F. (2005) ‘Aspects of effective mesoscale, short-range ensemble forecasting’, *Weather and Forecasting*, 20(3), pp. 328–350. doi: 10.1175/WAF843.1.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., and Johnson, N. A. (2008) ‘Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds’, *TEST*, 17(2), pp. 211–235. doi: 10.1007/s11749-008-0114-x.

Lindgren, F. and Rue, H. (2015) ‘Bayesian spatial modelling with R-INLA’, *Journal of Statistical Software*, 63, pp. 1–25. doi: 10.18637/jss.v063.i19.

Liu, C. and Rubin, D. B. (1994) ‘The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence’, *Biometrika*, 81(4), pp. 633–648. doi: 10.2307/2337067.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005) ‘Using Bayesian model averaging to calibrate forecast ensembles’, *Monthly Weather Review*, 133(5), pp. 1155–1174. doi: 10.1175/MWR2906.1.

Sloughter, J. M., Gneiting, T., and Raftery, A. E. (2010), ’Probabilistic wind speed forecasting using ensembles and Bayesian model averaging’, *Journal of the American Statistical Association*, 105(489), pp. 25-35.

Wilks, D. S. (1990) ‘Maximum likelihood estimation for the gamma distribution using data containing zeros’, *Journal of Climate*, 3(12), pp. 1495–1501. https://doi.org/10.1175/1520-0442(1990)003<1495:MLEFTG>2.0.CO;2

Wood, S. N. and Augustin, N. H. (2002) ‘GAMs with integrated model selection using penalized regression splines and applications to environmental modelling’, *Ecological Modelling*, 157(2), pp. 157–177. doi: 10.1016/S0304-3800(02)00193-X.