Inflow Forecasting for Hydropower Operations: Bayesian Model Averaging for Postprocessing Hydrological Ensembles

Andreas Kleiven[1,2] and Ingelin Steinsland[1]

[1] Department of Mathematical Sciences
Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
[2] Department of Industrial Economics and Technology Management
Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

Abstract. This paper contributes to forecasting of renewable infeed for use in dispatch scheduling and power systems analysis. Ensemble predictions are commonly used to assess the uncertainty of a future weather event, but they often are biased and have too small variance. Reliable forecasts for future inflow are important for hydropower operation, and the main purpose of this work is to develop methods to generate better calibrated and sharper probabilistic forecasts for inflow. We propose to extend Bayesian model averaging with a varying coefficient regression model to better respect changing weather patterns. We report on results from a case study from a catchment upstream of a Norwegian power plant during the period from 24 June 2014 to 22 June 2015.

Keywords: Bayesian model averaging, probabilistic postprocessing, inflow forecasting.

1 Introduction

Hydrological forecasting plays an important role in a variety of applications, ranging from flood prevention to water resource management and hydropower production. Forecasting inflow to hydropower reservoirs for operation and scheduling is the focus of this work. Future streamflows are uncertain, and forecasts generated from hydrological models are subject to errors. In order to quantify the uncertainty of future streamflows, it is common to generate an ensemble of forecasts with perturbations made for both the initial state and the model formulation for each member of the ensemble. The resulting ensemble can be interpreted as a probabilistic forecast. However, the ensemble forecasts tend to be underdispersive, meaning that the observed value too often lies outside the ensemble range. Therefore, statistical postprocessing methods are essential in order to obtain calibrated and sharp probabilistic forecasts.

A widely used postprocessing methodology for ensemble forecasts is Bayesian model averaging (BMA) [1]. In the BMA methodology, a component probability density function (pdf) is assigned to each ensemble member forecast, and the BMA probabilistic forecast is given by a weighted average of the individual ensemble member pdfs. Another popular postprocessing method is Ensemble model output statistics (EMOS) [2]. This method is based on multiple linear regression. An advantage with the BMA methodology is that the method respects the dynamics in the ensemble.

In the original BMA approach for postprocessing of forecast ensembles, a Gaussian pdf is assigned to the ensemble members [1]. Extensions of the BMA methodology have been developed for cases where the dependent variable deviates from the Gaussian case. Sloughter et al. [3] modified the method to apply to precipitation forecasts by introducing a discrete-continuous model which combines a logistic regression model and gamma distributions. Moreover, BMA using gamma distributions as the component pdfs has been applied to wind speed forecasting [4]. Furthermore, Duan et al. [5] used the BMA approach to generate probabilistic hydrological forecasts after transforming streamflow values using the Box-Cox transformation.

In this paper we aim to generate reliable probabilistic forecasts for inflow by extending the original BMA methodology. Many stochastic optimization methods used for operational purposes often require a large number of inflow scenarios as input, and inflow forecasts in the form of predictive distributions are useful in the sense that one easily can generate many inflow scenarios from sampling. Séguin et al. [6] propose a method for the natural next step of our analysis, which is a transition from a probabilistic forecast to a scenario tree or a lattice, something that is useful for input in short-term hydropower operational optimization methods. We propose to extend the BMA methodology for ensemble forecasts with varying coefficient regression (VCR) [7]. We demonstrate the method in a case study from a catchment upstream of a Norwegian power plant during the period from 24 June 2014 to 22 June 2015.

2 Bayesian Model Averaging using Varying Coefficient Regression

The use of BMA for statistical postprocessing of forecast ensembles was introduced by Raftery et al. [1]. The BMA approach generates a probabilistic forecast in the form of a predictive pdf by combining deterministic forecasts from different models. We suggest to extend the BMA methodology by using a VCR model, which we denote BMA-VCR. The models presented below can be applied to each lead time individually, where lead time refers to the forecast horizon.

We assume that the ensemble members are exchangeable, meaning that they are treated equally. Therefore, we present the BMA methodology for exchangeable member forecast. First, we follow the approach of Raftery et al. [1] and consider the normal distribution with mean $\alpha + \beta x_m$ and standard deviation τ

as the ensemble member pdfs. The BMA probabilistic forecast is then given by

$$\begin{aligned} &f(y|x_1,...,x_M) = \frac{1}{M}\sum_{m=1}^{M} g(y|x_m) \\ &Y|x_m \sim \mathcal{N}(\mu_m, \tau^2) \\ &\mu_m = \alpha + \beta x_m, \end{aligned} \tag{1}$$

where x_m is the deterministic forecast from ensemble member m, Y is the random variable representing future inflow to be forecasted, and M is the size of the ensemble. The bias-correction parameters, α and β, are equal for each ensemble member and the weights for exchangeable member forecasts are $\frac{1}{M}$. However, such a simple linear bias-correction does in general not provide good predictions for heteroskedastic and non-Gaussian model errors, which is likely to occur in hydrological forecasting [8]. To easier incorporate local weather patterns, we suggest to apply a nonlinear bias-correction in the form of a VCR model. VCR models are a class of generalized linear regression models where the coefficients are allowed to vary as functions of other variables. We let the BMA bias-correction parameters vary throughout time t, and a VCR model can then be described by

$$\begin{aligned} &\alpha_t = \alpha_{t-1} + a_t, \qquad a_t \sim N(0, \delta^{-1}) \\ &\beta_t = \beta_{t-1} + b_t, \qquad b_t \sim N(0, \delta^{-1}), \end{aligned} \tag{2}$$

where we restrict the precision parameter δ to be equal for both processes. We refer to δ as a precision parameter since the larger value, the less variance. In the BMA-VCR model, we include both static bias-correction parameters α and β and dynamic parameters α_t and β_t, which leads to the following form of the BMA-VCR probabilistic forecast

$$\begin{aligned} &f(y|x_1,...,x_M) = \frac{1}{M}\sum_{m=1}^{M} g(y|x_m) \\ &Y|x_m \sim \mathcal{N}(\mu_m, \tau^2) \\ &\mu_m = (\alpha + \alpha_t) + (\beta + \beta_t)x_m \\ &\alpha_t = \alpha_{t-1} + a_t, \quad a_t \sim N(0, \delta^{-1}) \\ &\beta_t = \beta_{t-1} + b_t, \quad b_t \sim N(0, \delta^{-1}). \end{aligned} \tag{3}$$

The static parameters α and β represent the total bias between forecast and observation pairs from a training period, and the dynamic parameters α_t and β_t evolve from time $t = 1$. The parameter δ decides the flexibility of the dynamic parameters. A small δ gives a good fit to training data, but generally not good predictions, i.e. overfitting, while a large value of δ gives less flexibility for the dynamic parameters. By letting $\delta^{-1} = 0$ the BMA-VCR model formulation in (3) coincide with the original BMA model defined in (1). The assumption that δ^{-1} is identical for α_t and β_t is reasonable as the dependency between the estimators for α_t and β_t is high. If $\alpha + \alpha_t = 0$, and $\beta + \beta_t = 1$, the mean of an ensemble member forecast, μ_m, will be the deterministic forecast x_m. If $\beta + \beta_t < 1$, we expect $\alpha + \alpha_t > 0$. Furthermore, if $\beta + \beta_t > 1$, we expect $\alpha + \alpha_t < 0$.

3 Parameter Estimation

In the original BMA methodology for ensemble forecasts, a sliding window of constant size, consisting of forecast and observation data from the most recent history, is used to train the model. In the VCR models, there are dynamic parameters that evolve from time $t = 1$. We use Bayesian inference for estimation of bias-correction parameters $\alpha, \beta, \alpha_t, \beta_t$. The variance parameter τ is then estimated from a sliding window training period in the same way as in the work of Raftery et al. [1]. The last model parameter, the precision parameter δ, is estimated based on predictive performance. For inference, we apply integrated nested Laplace approximations (INLA) [9, 10]. INLA is a method for performing approximate Bayesian inference. As an alternative to simulation-based Monte Carlo integration, INLA uses the analytic approximation with the Laplace method, which leads to computational benefits. Furthermore, R-INLA [11], which is an open source software, is suitable for parameter estimation in the BMA-VCR model.

4 Forecast Verification

Probabilistic forecasts take the form of predictive pdfs, and in order for the forecast to be useful, it is important to assess the predictive performance. The models are evaluated according to calibration and sharpness. Calibration is the statistical consistency between the predictive pdfs and the corresponding observed values. Sharpness is a measure of uncertainty of the predictive pdfs.

The verification rank histogram (VRH) is often used to assess calibration of ensemble forecasts [1, 3, 12]. The VRH is computed by arranging the ensemble forecasts and the corresponding observation in increasing order. To assess calibration of probabilistic forecasts, the probability integral transform (PIT) is common to apply [1, 3, 13]. The probabilistic forecast is calibrated if the PIT values, which is the value of the predictive cdf at the corresponding observation, are uniformly distributed. Uniformity can be assessed by making a histogram of PIT values. The shape of the VRH and the PIT histogram, gives an indication whether the probabilistic forecast is calibrated. Hump-shaped histograms indicate that the probabilistic forecast is overdispersed, which means that the prediction intervals on average are too wide. U-shaped histograms indicates underdispersion, meaning that the prediction intervals on average are too narrow. Asymmetrical histograms occur when the probabilistic forecast is biased.

Proper scoring rules are often used to assess the predictive performance of a probabilistic forecast. A scoring rule is proper if the expected score is minimized when the issued forecast is the true distribution of the quantity to be forecasted [14]. The continuous ranked probability score (CPRS) is a proper scoring rule that measures both calibration and sharpness of a probabilistic forecast [15]. The

CRPS measures the difference between the predicted and occurred cumulative distributions. The value of the CRPS is non-negative and the smaller value the better quality of the probabilistic forecast. For deterministic forecasts, the CRPS reduces to the absolute error, hence it is possible to compare the performance of probabilistic forecasts and deterministic forecasts.

5 Data and Study Area

The ensemble forecasts used in this study are generated from the Hydrologiska Byråns Vattenbalansavdelning (HBV) model [16]. The model has a number of free hydrological parameters that are estimated from training data, and the start state is estimated using observed precipitation and temperature from the history. Ensembles of temperature and precipitation forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) are used as input in the HBV-model. The ensemble size in this study is $M = 51$, and the ensemble forecasts are treated equally, i.e. they are exchangeable.

In the case study, we consider the Osali catchment which is a part of the Ulla-Førre hydropower complex south west in Norway [17, 18]. Daily inflow observations, in unit $m^3\,s^{-1}$, are recorded and data are provided by Statkraft, which is the largest hydropower producer in Norway. The method is evaluated for lead time $l = 1$ day, where lead time refers to the forecast horizon.

6 Results and Discussion

We apply the BMA-VCR method to inflow forecasting from the Osali catchment. The method is tested in the period from 24 June 2014 to 22 June 2015. We analyze how the precision parameter δ influence the predictive performance by considering mean CRPS, which is the average CRPS taken over all days in the period under study. The lower mean CRPS, the better predictive performance. Calibration is assessed through the PIT histogram.

Mean CRPS is given as a function of the inverse precision parameter δ^{-1} in Fig. 1. The original BMA method is the purple horizontal line, and the mean CRPS of the ensemble is given by the green line. The BMA-VCR method, which corresponds to a non-linear bias-correction in the original BMA methodology is shown in red. Where lines intersect means that the predictive performance is equally good. We observe that a large value of δ^{-1} leads to large mean CRPS, which in this case corresponds to overfitting and poor predictive performance. We observe that an inverse precision parameter value close to 0.11 provides a good forecasting performance for the BMA-VCR method. We get mean CRPS values 0.57, 0.47, and 0.39 for the raw ensemble, BMA method, and BMA-VCR method respectively.

We observe from Fig. 1 that the right choice for δ^{-1} is important. Values between $\delta^{-1} = 0.07$ and $\delta^{-1} = 0.13$ leads to better predictive performance compared to the original BMA method. Values outside this interval leads higher

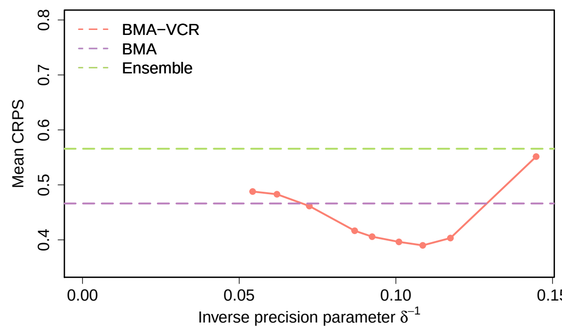

Fig. 1: Mean CRPS as a function of the inverse precision parameter δ^{-1}. The figure shows the potential of including a VCR model in the BMA methodology for postprocessing of hydrological ensembles. We observe that mean CRPS is lower for BMA-VCR compared to BMA with static parameters for certain values of δ^{-1}.

CRPS values. The computation time for estimating parameters in the BMA-VCR method, using Bayesian inference, is longer compared to the BMA method, which uses maximum likelihood estimation. The computation time increases with increasing amount of data used for fitting the model. With small amount of data available, the computation time of the BMA-VCR method is similar to the original BMA method. As more data become available, the estimation procedure takes longer time. For operational use, a sliding window with constant size can be applied to reduce computation time.

The PIT histogram obtained for the probabilistic forecast from the BMA-VCR method and the VRH for the ensemble forecasts are provided in Fig. 2. The horizontal dotted line indicate the height of the bars for a perfectly calibrated forecast. We observe that the VRH for lead time 1 is strongly u-shaped and slightly biased. This means that the ensemble underestimate variance. The PIT histogram obtained from the probabilistic forecast of the BMA-VCR method is closer to uniform.

The method can be further extended and applied to a higher-dimensional system, but this is not tested in this work. For multiple catchments, the dependency between corresponding ensemble members will be handled by the current method. However, further extensions to the proposed methodology are needed.

7 Conclusion

In this work we have presented a new postprocessing method for hydrological ensembles. We have suggested to extend the original BMA approach for post-processing of ensemble forecasts with a VCR model. The performance of the postprocessing methods was demonstrated in a case study of the Osali catch-

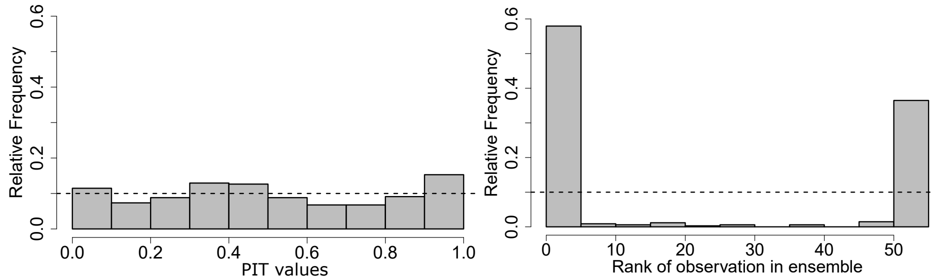

Fig. 2: he PIT histogram obtained from the probabilistic forecast of the BMA-VCR method (left) and the VRH from the raw ensemble (right)

ment in the south-western part of Norway for lead time $l = 1$ day. The results showed that applying a non-linear regression for the bias-correction parameters in the original BMA methodology has great potential to improve the predictive performance of hydrological ensembles for short lead times.

Acknowledgements. The authors would like to thank Stein-Erik Fleten for valuable discussions. We also thank Statkraft and Stian Solvang Johansen for providing data and motivation behind this work. Andreas Kleiven acknowledges support from the Research Council of Norway, through HydroCen, project number 257588. For Ingelin Steinsland this project was supported by the Research Council of Norway, project 250362.

References

1. Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). *Using Bayesian model averaging to calibrate forecast ensembles.* Monthly Weather Review, 133(5), 1155-1174.
2. Gneiting, T., Raftery, A. E., Westveld III, A. H., & Goldman, T. (2005). *Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation.* Monthly Weather Review, 133(5), 1098-1118.
3. Sloughter, J. M. L., Raftery, A. E., Gneiting, T., & Fraley, C. (2007). *Probabilistic quantitative precipitation forecasting using Bayesian model averaging.* Monthly Weather Review, 135(9), 3209-3220.
4. Sloughter, J. M., Gneiting, T., & Raftery, A. E. (2010). *Probabilistic wind speed forecasting using ensembles and Bayesian model averaging.* Journal of the american statistical association, 105(489), 25-35.
5. Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). *Multi-model ensemble hydrologic prediction using Bayesian model averaging.* Advances in Water Resources, 30(5), 1371-1386.
6. Séguin, S., Fleten, S. E., Côté, P., Pichler, A., & Audet, C. (2017). *Stochastic short-term hydropower planning with inflow scenario trees.* European Journal of Operational Research, 259(3), 1156-1168.

7. Hastie, T., & Tibshirani, R. (1993). *Varying-coefficient models.* Journal of the Royal Statistical Society. Series B (Methodological), 757-796.
8. Vrugt, J. A., & Robinson, B. A. (2007). *Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging.* Water Resources Research, 43(1).
9. Rue, H., Martino, S., & Chopin, N. (2009). *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.* Journal of the Royal Statistical Society: Series B (statistical methodology), 71(2), 319-392.
10. Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA.* John Wiley & Sons.
11. Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). *Bayesian computing with INLA: new features.* Computational Statistics & Data Analysis, 67, 68-83.
12. Hamill, T. M. (2001). *Interpretation of rank histograms for verifying ensemble forecasts.* Monthly Weather Review, 129(3), 550-560.
13. Dawid, A. P. (1984). *Present position and potential developments: Some personal views: Statistical theory: The prequential approach.* Journal of the Royal Statistical Society. Series A (General), 278-292.
14. Gneiting, T., & Raftery, A. E. (2007). *Strictly proper scoring rules, prediction, and estimation.* Journal of the American Statistical Association, 102(477), 359-378.
15. Hersbach, H. (2000). *Decomposition of the continuous ranked probability score for ensemble prediction systems.* Weather and Forecasting, 15(5), 559-570.
16. Bergström, S. (1992). *The HBV model: Its structure and applications.* Swedish Meteorological and Hydrological Institute.
17. Engeland, K., & Steinsland, I. (2014). *Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times.* Water Resources Research, 50(1), 182-197.
18. Engeland, K., Steinsland, I., Johansen, S. S., Petersen-Øverleir, A., & Kolberg, S. (2016). *Effects of uncertainties in hydrological modelling. A case study of a mountainous catchment in Southern Norway.* Journal of Hydrology, 536, 147-160.