Final Version Published

Ensemble Models

Montserrat Fuentes
Statistics Department, North Carolina State University, Raleigh, NC, USA
Kristen Foley
National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

Abstract
An ensemble, or sample, of competing numerical models has been used in many applications to represent different predictions of the true state of a physical system. Ensembles of computer models (e.g. weather, climate, air quality, ocean models, etc.) are often used to forecast future states of a physical system and to quantify uncertainty in the numerical model predictions. Various statistical methods have been proposed to improve ensemble predictions from deterministic computer simulations based on actual measurements of the physical systems (e.g. data assimilation, Bayesian model averaging). Ensemble data mining methods have also been developed for a wide variety of applications to combine different versions of a statistical model (e.g. time series models, simple regression models, neural networks, etc.) to improve the predictive model performance. We present different statistical criteria that have been proposed to select or weight ensemble members for both numerical model-based and statistical model-based ensembles.
Keywords: data assimilation, ensemble Kalman filter, Bayesian Model Averaging, ensemble data mining, bagging, boosting

Ensembles of Computer Models

Ensemble forecasting has been used for operational numerical weather prediction in the US and Europe since the early 1990s. An ensemble of weather or climate forecasts is used to characterize the two main sources of uncertainty in computer models of physical systems: (1) initial conditions and forcing variables (i.e. parametric uncertainty); and (2) imperfections in the formulation of the physical model, such as those due mathematical approximations or lack of understanding about the fundamental mechanisms underlying the physical process (i.e. structural uncertainty) (Pinder et al. 2009). Ensemble models have also been used for improving prediction of and quantifying uncertainty in models of other physical systems such as air quality modeling, ocean modeling, land surface modeling and hydrological modeling. Ensemble models may be created by using simulations from different physical models that have been developed independently by different research organizations. Alternatively, ensemble models may be propagated by using different formulations of a single modeling system. In this case, multiple model simulations are generated by perturbing initial conditions (e.g. based on breeding or singular vectors), using different forcing fields (e.g. different boundary conditions or other uncertain model inputs), or using competing (but scientifically defensible) parameterizations of a single process within the system. The advantage of using ensemble models is that the ensemble mean prediction is typically more accurate than the predictions from the individual ensemble members. More importantly, the spread of the ensemble provides an estimate of the reliability of the ensemble prediction which can vary over time and space (Kalnay 2003).

In the geophysical sciences the modeled state vector that describes the state of the atmosphere or other physical system is usually of very high dimension (e.g. 10^3-10^6) and integrating the model forward in time can require a great deal of time and computational resources. As a result, only a relatively small number of simulations of computer models are often feasible (e.g. 10 -100). Thus ensemble members must be carefully selected and/or weighted in order to adequately characterize uncertainty in the model output. Many statistical approaches have been developed to address these problems by combining numerical model output with observations. In the field of climate modeling, it is often unclear how to utilize observational data that have very different spatial and temporal coverage from the model output. In this case, hierarchical statistical models have been applied to combine climate model output from a limited number of model runs to better characterize the distribution of the state including cross-dependencies between different state variables (e.g. Sain et al. 2011).

Combining Model Output and Observations

Statistical methods have been used to enhance ensemble models by combining model and observed information to create more accurate initial conditions for a forecast or prediction cycle. This process is a type of state space estimation problem and is referred to as data assimilation. A wide variety of ensemble data assimilation methods have been developed in the last twenty years. In particular, the ensemble Kalman Filter (EnKF, Eversen 2003) uses an optimal weight matrix (i.e. the Kalman gain matrix) to update the forecast or "first guess" of the state vector based on the difference between the observed values and the first guess. The weight matrix is a function of the forecast error covariance and the observation error covariance. With careful treatment of the sample forecast covariance structure, the EnKF has been successfully applied even when the ensemble size is much smaller than the dimension of the state vector (Houtekamer and Mitchell 2001; Hamill et al. 2001; Bickel and Levina 2008). A limitation of the EnKF is that the underlying Kalman filtering algorithm is based on the assumption that the system being modeled is linear with Gaussian errors. The particle filter improves on the Monte Carlo approach applied in the EnKF by utilizing an importance sampling framework. The particle filter is applicable to highly non-Gaussian probability distributions, but thus far has only been implemented in relatively low dimensional numerical models (see Snyder et al. 2008 for further discussion on the use of particle filters for geophysical systems). Bocquet et al. (2010) present a review of several advanced non-Gaussian data assimilation methods, including the particle filter.

Another approach for combining an ensemble of model predictions and observed data is to post-process the ensemble of model runs based on past observations and past simulations. Bayesian model averaging (BMA) can be used to estimate a set of weights for the ensemble members based on how closely they match observations over a given training period (Raftery et al. 2005). The BMA predictive probability density functions are more accurate and better calibrated than the original ensemble and the estimated weights are used to evaluate the usefulness of individual ensemble members. Recent extensions of the BMA approach include using non-Gaussian component distributions (Sloughter et al. 2007), incorporating spatially correlated error fields (Berrocal, Raftery, and Tilmann 2007), and applying Markov Chain Monte Carlo (MCMC) estimation techniques (Vrugt, Kiks, and Clark 2008).

Ensemble Data Mining Methods

Ensemble data mining methods are machine learning methods that utilize information from multiple statistical models to improve the predictive model performance compared to the performance of any one model (Oza 2006). Since the mid-1990s, much has been published in machine learning literature on how to train a set of base models and then combine or weight predictions from competing models. Two common ensemble methods for classification machine learning problems are Bagging (Breiman 1994) and Boosting (Freund and Schapire 1996). The key feature of these methods is that they promote diversity across ensemble members by training each base model with a different subset of input data or, in the case of Boosting, with different sets of weights applied to the input data. In Bagging, the ensemble member predictions are combined using simple majority or plurality voting techniques. Bagged ensembles have been shown to improve upon their base model predictions if differences in the training sets create significant differences in the models (Breiman 1994). In Boosting, the base models may be combined by simple averaging or weighted averaging based on the probability values for each base model produced by the algorithm. Different weighting methods have been proposed, e.g. Mixtures of Experts (Jordan and Jacobs 1994; Titsias and Likas 2002), and Principle Component Regression (Merz and Pazzani 1999). Boosting algorithms have also performed well in practice, however performance has been shown to degrade when the training data is noisy (Dietterich 2000).

Disclaimer

Although this manuscript has been reviewed and approved for publication, it does not necessarily reflect the policy or views of the US Environmental Protection Agency.

REFERENCES

Berrocal, V.J., Raftery, A.E., Gneiting, T. Combining spatial statistical and ensemble information in probabilistic weather forecasts. Monthly Weather Review, 135, 1386-1402, 2007.

Bickel, P., Levina, E. Regularized estimation of large covariance matrices. Annals of Statistics, 36, 199-227, 2008.

Bocquet, M., Pires, C., Wu, L. Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation, Monthly Weather Review, 138, 2997 – 3023, 2010.

Breiman, L. Bagging predictors, Technical Report 421, Department of Statistics, University of California, Berkely, 1994.

Dietterich, T. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization, Machine Learning, 40, 139 – 158, 2000.

Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dynamics, 53, 343-367, 2003.

Freund, Y., Schapire, R. Experiments with a new Boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning, 148 – 156. Morgan Kaufmann, 1996.

Hamill, T.M., Whitaker, J.S., Snyder, C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review, 129, 2776-2790, 2001.

Houtekamer, P.L., Mitchell, H.L. A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review, 129, 123-137, 2001.

Jordan, M.I., Jacobs, R.A. Hierarchical mixture of experts and the EM algorithm. Neural Computation, 6, 181-214, 1994.

Kalnay, E. Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, UK, 2003.

Merz, C.J., Pazzani, M.J. A principal component approach to combining regression estimates. Machine Learning, 26, 9 – 32, 1999.

Pinder, R.W., Gilliam, R.C., Appel, K.W., Napelenok, S.L., Foley, K.M., Gilliand, A.B. Efficient probabilistic estimates of surface ozone concentrations using an ensemble of model configurations and direct sensitivity calculations. Environmental Science and Technology, 43, 2388-2393, 2009.

Oza, N.C. Ensemble Data Mining Methods. In Encyclopedia of Data Warehousing and Mining, 448 – 453, Idea Group Inc., 2006.

Raftery, A. E., Gneiting, T., Balabdaoui, F., Polakowski, M. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Monthly Weather Review, 133, 1155-1174, 2005.

Sain, S.R., Furrer, R., Cressie, N. A spatial analysis of multivariate output from regional climate models. The Annals of Applied Statistics, 5, 150-175, 2011.

Sloughter, J.M., Raftery, A.E., Gneiting, T., Fraley, C. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. Monthly Weather Review, 135, 3209-3220, 2007.

Snyder, C., Bengtsson, T., Bickel, P., Anderson, J. Obstacles to high-dimensional particle filtering. Monthly Weather Review, 136, 4629-4640, 2008.

Titsias, M.K., Likas, A. Mixture of experts classification using a hierarchical mixture model. Neural Computation, 14, 2221 – 2244, 2002.

Vrugt, J.A., Diks, C.G.H., Clark, M.P. Ensemble Bayesian model averaging using Markov Chain Monte Carlo sampling. Environmental Fluid Mechanics, 579-595, 2008.