Entropy Ensemble Filter: Does information content assessment of bootstrapped training datasets before model training lead to better trade-off between ensemble size and predictive performance?

Hossein Foroozand and Steven V. Weijs

Civil Engineering Dept., University of British Columbia

Abstract

Machine learning is the fast-growing branch of data-driven models, and its main objective is to use computational methods to become more accurate in predicting outcomes without being explicitly programmed. In this field, a way to improve model predictions is to use a large collection of models (called ensemble) instead of a single one. Each model is then trained on slightly different samples of the original data, and their predictions are averaged. This is called bootstrap aggregating, or Bagging, and is widely applied. A repeated question in previous works was: how to determine the bagging ensemble size of training data sets for tuning the weight in machine learning? The computational cost of ensemble-based methods scales with the size of the ensemble, but excessively reducing the ensemble size comes at the cost of reduced predictive performance. The choice of ensemble size was often determined based on the size of input data and available computational power, which can become a limiting factor for larger datasets and complex models' training. In this research, it is our hypothesis that if an ensemble of artificial neural networks (ANN) models or any other machine learning technique uses the most informative ensemble members for training purpose rather than all bootstrapped ensemble members, it could reduce the computational time substantially without negatively affecting the performance of simulation.

Introduction

Advanced computational methods, including artificial neural networks (ANN), process input data in the context of previous training history on a defined sample database to produce relevant output. To avoid negative effects of over-fitting, an ensemble of models is sometimes used in prediction. Bagging (abbreviated from Bootstrap AGGregatING) [1] developed from the idea of bootstrapping [2] in statistics. Despite its common application, the bagging method is considered to be computationally expensive, particularly when used to create new training data sets out of large volumes of observations [3–4].

In this poster, we combine materials from the Entropy Ensemble Filter (EEF) method [5] and its first real-world application [6] to highlight the method's advantages and limitations. Entropy can be defined as uncertainty of a random variable or, conversely, the information that samples of that random variable provide. In this work, entropy is used as a measure of information content for each bootstrap resample of the dataset. The method selects high entropy bootstrap samples for ensemble model training, aiming to maximize information content in the selected ensemble. We applied our proposed method on a simulation of synthetic data with the ANN machine learning technique. Also, its application is tested in forecasting the tropical Pacific sea surface temperatures (SST) anomalies based on the neural-network forecast model proposed by Wu et al. [7].

Methodology and Data

The philosophy of the EEF method [5] is rooted in using self-information of a random variable, defined by Shannon's information theory [8] for selection, to provide direction in the inherent randomness of ensemble models which are created by bootstrapping. In this work, the focus is on selecting an ensemble of training datasets before model training (Fig1).

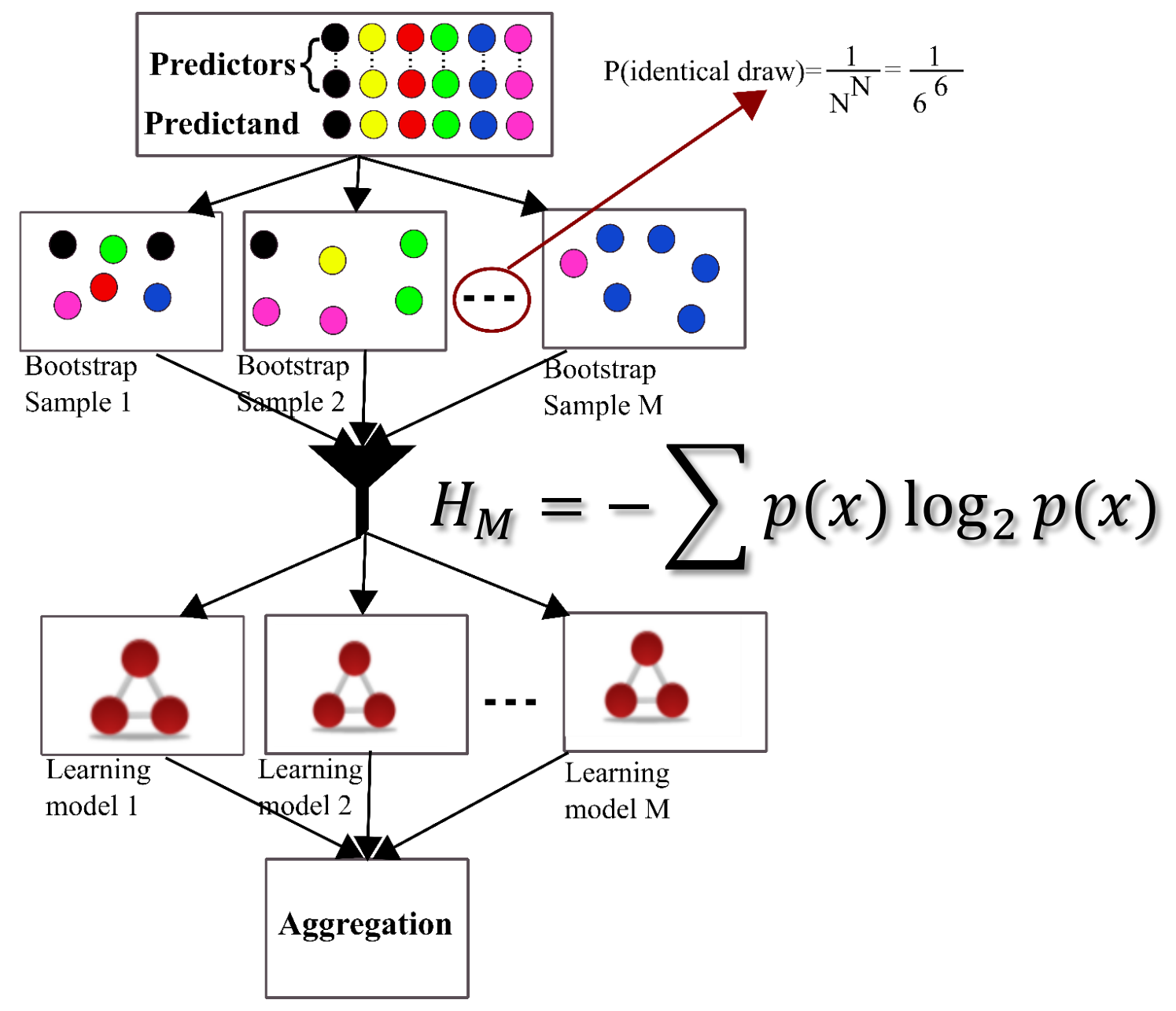


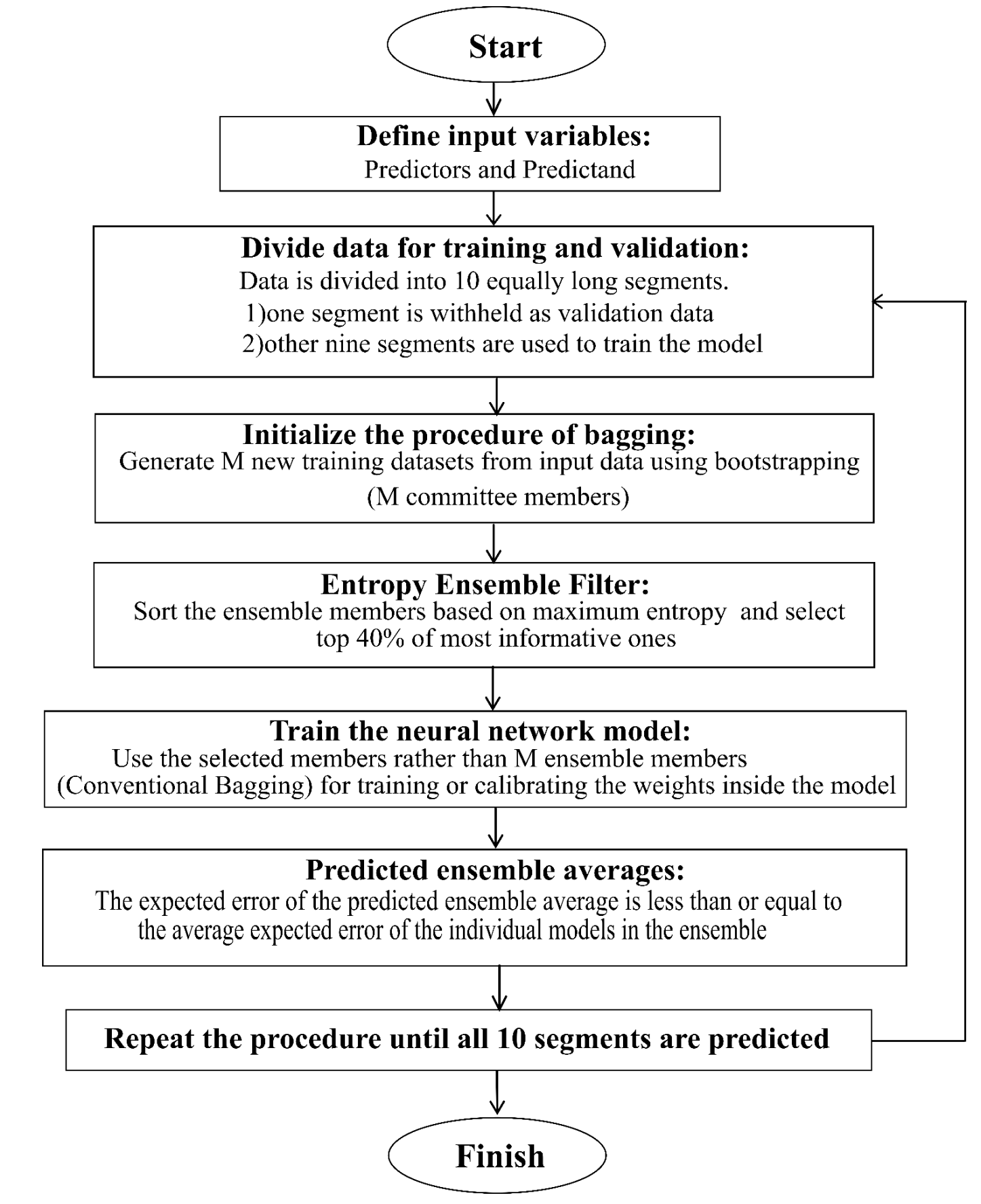
Figure 1: How modified Bagging works?

Synthetic Data Simulation [5]:

We use artificial sinusoidal signal that we corrupt with noise before model training to examine the model's capability to capture the essence of the signal from the noisy signal. The noisy signal was used as an input in the bagging procedure to generate an ensemble of input datasets, the chosen members by the EEF method are used for training ANN's and subsequently generating the simulation result for each member. The predictions were evaluated by calculating root mean square error (RMSE) against the target signal. To get insight in the trade-off between ensemble size (i.e., computation time) and accuracy in terms of RMSE, analysis of the error gradient with growing ensemble size was conducted (Fig3). In this analysis, the decrease in prediction error was compared between using the EEF method and conventional bagging with increasing ensemble size.

Sea Surface Temperatures (SST) Forecasts [6]:

The application of the EEF method is evaluated in neural network modeling of El Nino-southern oscillation. The goal is to forecast the first five principal components (PCs) of sea surface temperature monthly anomaly fields over tropical Pacific (Fig4), at different lead times from 3 to 15 months. The period 1979–2017. The model's structure developed by Wu et al. [7] is adopted, where sea level pressure (SLP) field and SST anomalies over Tropical Pacific were used to predict the five leading SST principal components at lead times from 3 to 15 months. The EEF method (Fig2) is applied in multiple linear regression (MLR) model and two neural network models, one using Bayesian regularization (labeled as BNN) and one Levenberg-Marquardt algorithm (labeled as NN) for training, and evaluate their performance. The conventional bagging uses 30 and 12 (labeled with no subscript and rand subscript, respectively) ensemble members for model training. In EEF method (labeled with subscript E), ensemble size is reduced to be 40% of the original one (Fig2).



Start

Define input variables:
Predictors and Predictand

Divide data for training and validation:
Data is divided into 10 equally long segments.
1)one segment is withheld as validation data
2)other nine segments are used to train the model

Initialize the procedure of bagging:
Generate M new training datasets from input data using bootstrapping
(M committee members)

Entropy Ensemble Filter:
Sort the ensemble members based on maximum entropy and select top 40% of most informative ones

Train the neural network model:
Use the selected members rather than M ensemble members (Conventional Bagging) for training or calibrating the weights inside the model

Predicted ensemble averages:
The expected error of the predicted ensemble average is less than or equal to the average expected error of the individual models in the ensemble

Repeat the procedure until all 10 segments are predicted

Finish

Figure 2: The flowchart of Entropy Ensemble Filter (EEF) method applied in the study [6].

Results

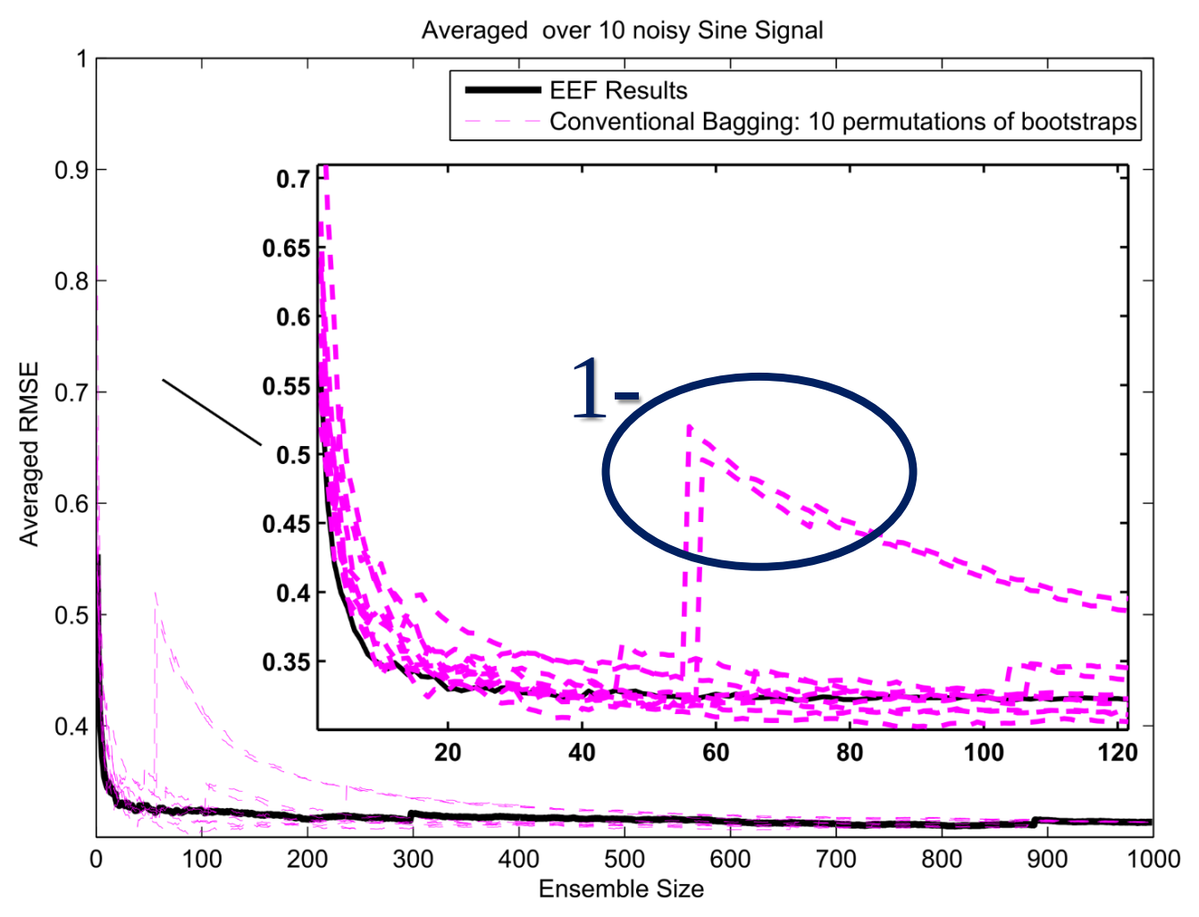


Figure 3: The error gradient analysis for sinusoidal signal and 1000 initial bootstrapped ensemble [5].

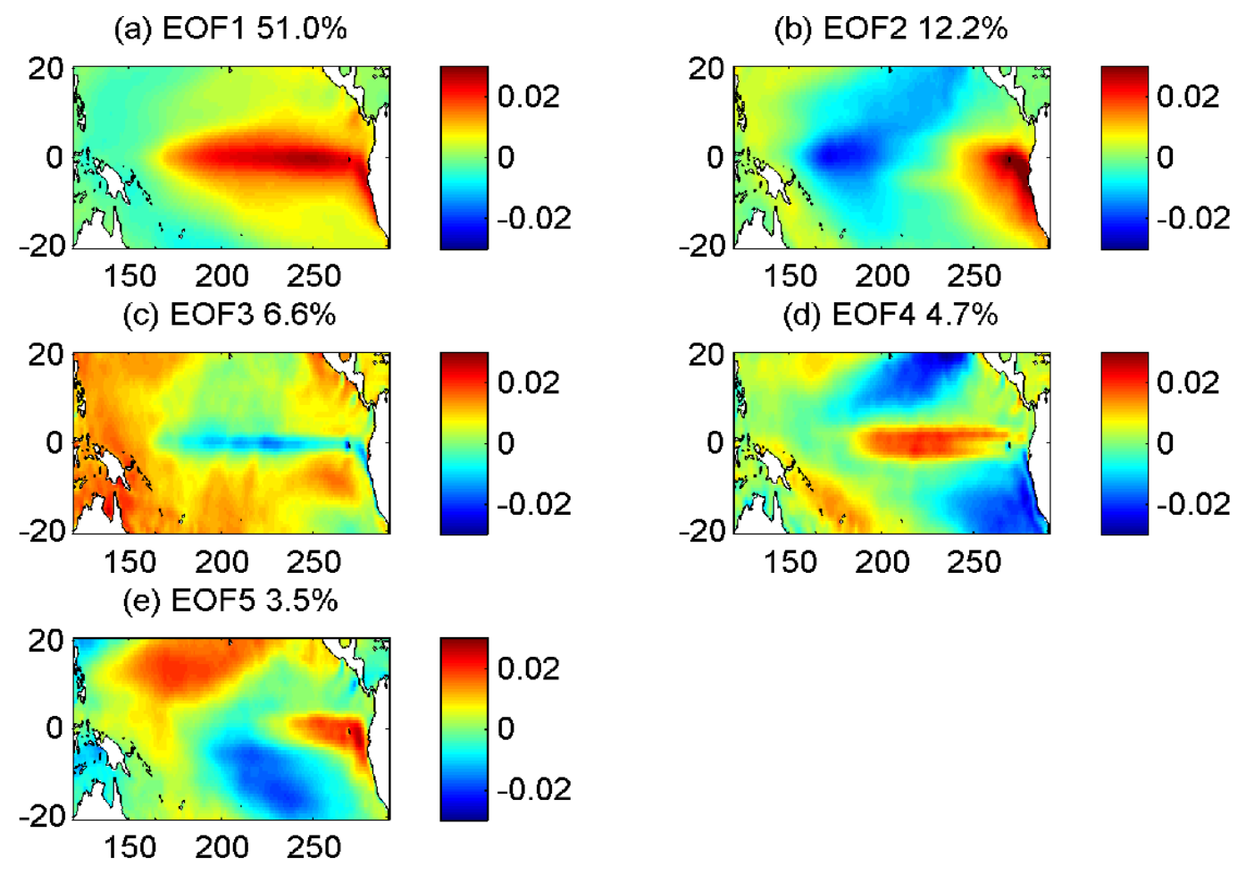


Figure 4: Spatial patterns of the first five PCA modes for the SST anomaly field [6].

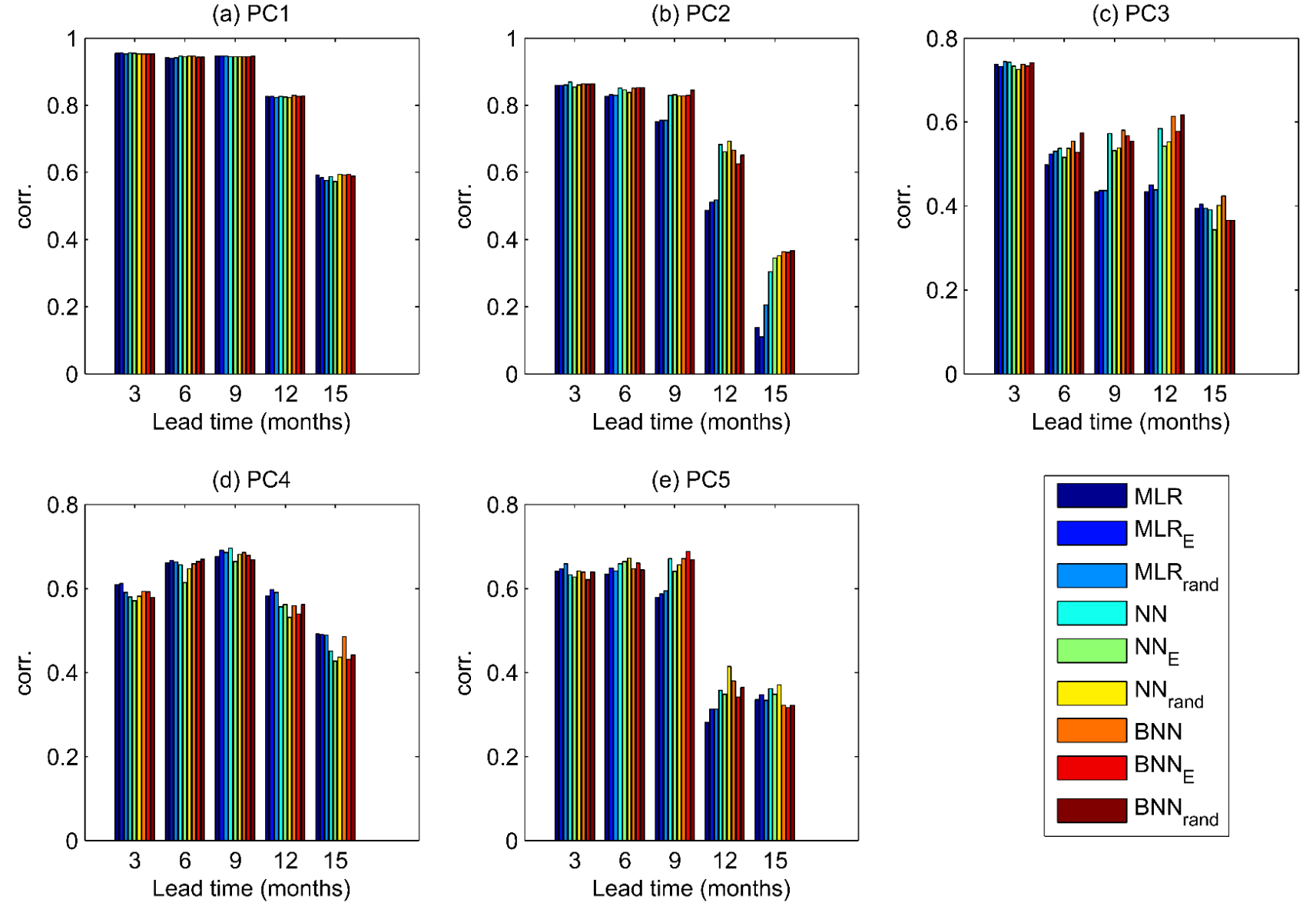


Figure 5: Correlation skill of predictions of the five leading principal components of the SST fields at lead times from 3 to 15 months [6].

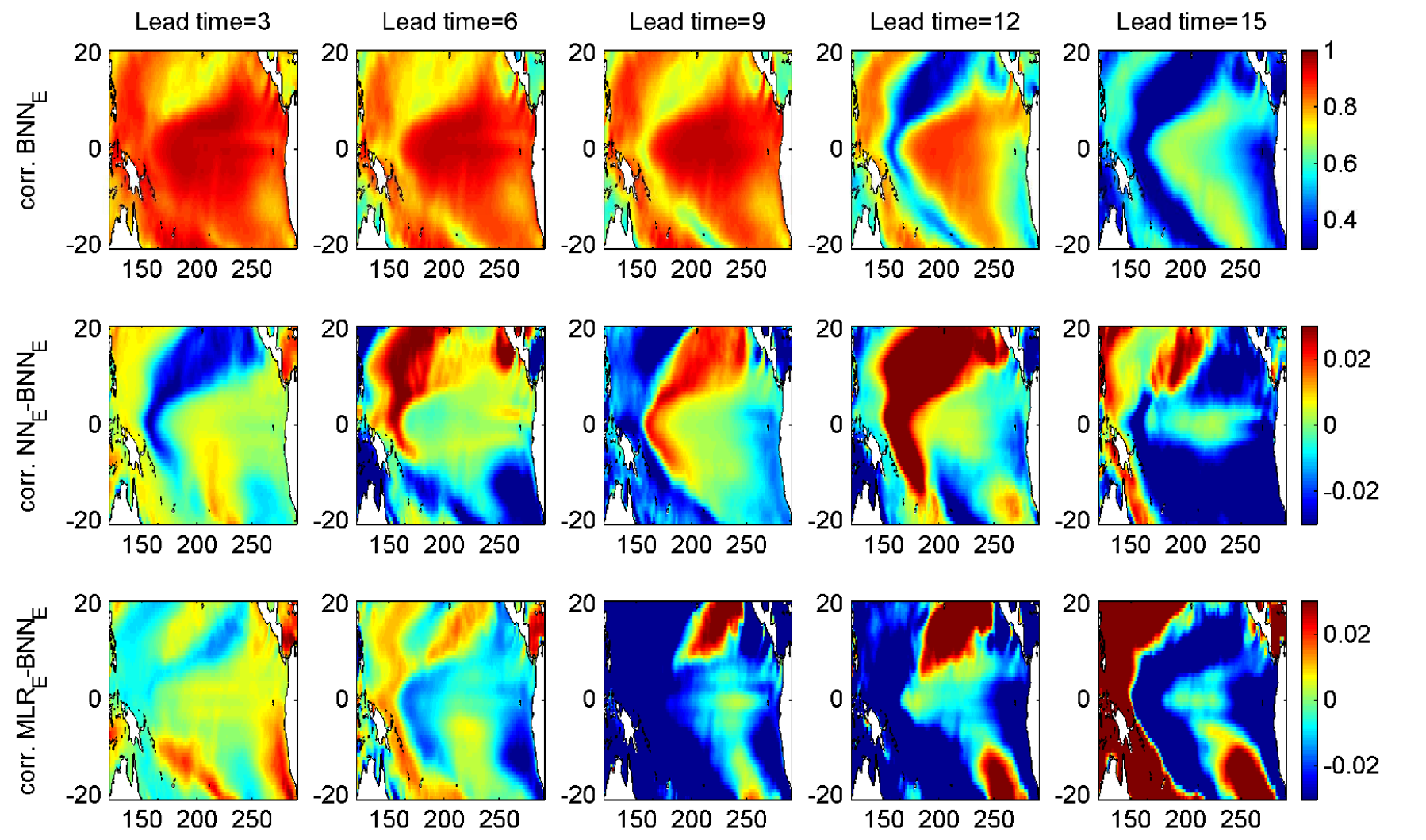


Figure 6: Forecast performance (correlation) per pixel of the forecast reconstructed from 5 PCs at lead times of 3–15. Top row: BNN_E model, middle and bottom rows: comparison of performance of NN_E and MLR_E over BNN_E [6].

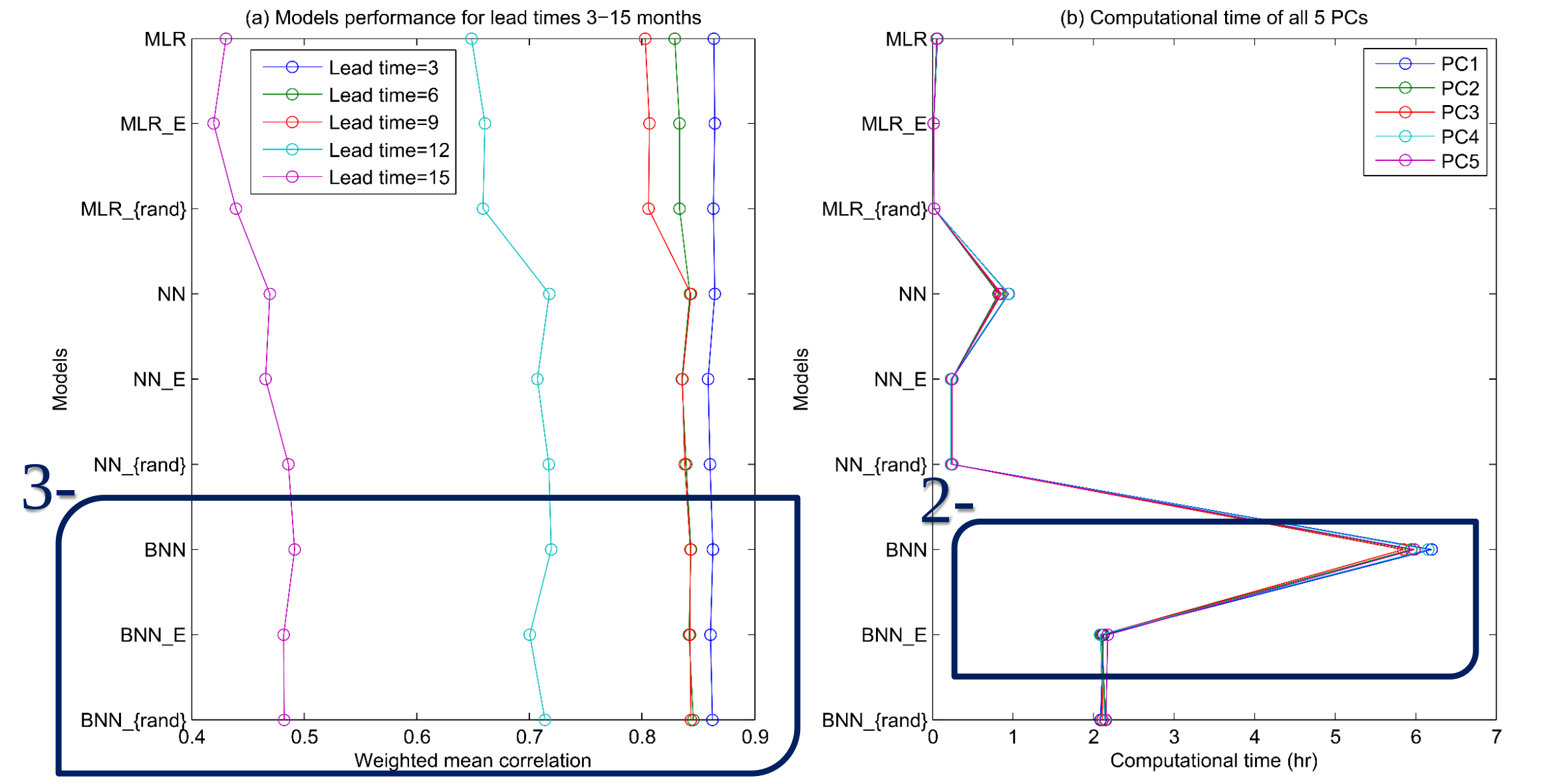


Figure7: Weighted mean correlation and computational time for all models [6].

Conclusion

1) The EEF method's idea of ranking and selecting the informative ensemble can lead to filter out outliers in large ensemble.
2) The EEF method can be useful to meet the computational power constraints for the continual arrival of new data, that necessitates frequent model updating. This method can make Bagging feasible for big datasets and complex model training.
3) For this particular case with small ensemble (30 samples), however, the conventional Bagging draws random ensemble that closely resemble the optimal ensemble from the EEF method. Thus, the neural network model with both bagging methods produced equally successful forecasts with the same computational efficiency.

References

1. Breiman, L (1996). Bagging Predictors. Mach. Learn. **Link**
2. Efron, B (1979). Bootstrap Methods: Another Look at the Jackknife. Ann. Stat. **Link**
3. Kasiviswanathan, K.S.; Sudheer, K.P (2013). Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. Stoch. Environ. Res. Risk Assess. **Link**
4. Wan, C.; Song, Y.; Xu, Z.; Yang, G.; Nielsen, A.H (2016). Probabilistic wind power forecasting with hybrid artificial neural networks. Electr. Power Compon. Syst. **Link**
5.Foroozand, H. and Weijs,S. V. (2017): Entropy Ensemble Filter:A Modified Bootstrap Aggregating (Bagging) Procedure to Improve Efficiency in Ensemble Model Simulation. Entropy J. **Link**
6. Foroozand, H., Radic, V., and Weijs, S. V. (2018): Application of Entropy Ensemble Filter in Neural Network Forecasts of Tropical Pacific Sea Surface Temperatures. Entropy J. **Link**
7.Wu, A.; Hsieh, W.W.; Tang, B (2006). Neural network forecasts of the tropical Pacific sea surface temperatures. Neural Netw. **Link**
8. Shannon, C (1948). A Mathematical Theory of Communication. Bell Syst. Tech. J. **Link**

CONTACT

Email: hosseinforoozand@yahoo.com
Email: steven.weijs@civil.ubc.ca