Journal of Hydrology 377 (2009) 284–299

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Combining single-value streamflow forecasts – A review and guidelines for selecting techniques

Dae Il Jeong [a,1], Young-Oh Kim [b,*]

[a] INRS-ETE, University of Québec, 490 de la Couronne, Québec, Canada G1K 9A9
[b] Department of Civil and Environmental Engineering, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea

ARTICLE INFO

Article history:
Received 25 February 2008
Received in revised form 22 April 2009
Accepted 25 August 2009

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Dong-Jun Seo, Associate Editor

Keywords:
Bias
Combining forecasts
Cross correlation
Forecast error variance
Hydrologic forecasting
Stationarity

SUMMARY

Choosing an appropriate method for combining single-value forecasts should depend on characteristics of the individual forecasts being combined and their relationships with each other. This study attempts to develop a guideline to choose effective combining techniques by using analytical derivations and/or hydrological experiments. The two most popular combining techniques, Simple Average (SA) and Weighted Average (WA), are compared from theoretical angles. The standard deviation of the combined forecast error is quantified as a function of the ratio of the standard deviation and the correlation coefficient between the two constituent forecast errors. Following the theoretical study, empirical research for eight combining methods including SA and WA methods was conducted to confirm the theoretical findings of this study and to verify results from other research carried out. The results of the empirical experiments are summarized to confirm the effects of the eight combining methods. The major findings include that: (1) SA yields reasonable results for any combination of forecasts when information of constituent forecasts is absent, (2) one cannot expect combining technique to yield significant improvement when two constituent forecasts are highly correlated, (3) the Regression and ANN combining methods can remove the effects of bias in the constituent forecasts and yield unbiased combining forecasts, and (4) when the constituent forecasts have nonstationary errors, a time-varying-weight combining method yields better results than the constant-weight methods in most cases. Based on these theoretical findings and empirical results, a guideline for combining methods is provided. The guideline suggests appropriate methods for combining single-value streamflow forecasts by considering bias and nonstationarity of the errors in the individual forecasts; the ratio of the error variance of any two forecasts and cross-correlation among the forecasts.

Introduction

Our lives face an endless series of imperfectly informed decisions: what to eat, whether or not to bring an umbrella, what move to make in a game, and numerous others. Any important decision is generally based on a wide variety of information, in particular in the form of *forecasts*.

As anyone else, investors want accurate forecasting. As a result, there may be dozens of professional analysts ready and willing to predict the price of any given stock. While some will believe any forecaster who has performed best in the past, the wise investor consults several good forecasters and then "combines" their forecasts with his/her own rule. The investor may consider numerous factors based on statistical analysis, anecdotal evidence, or personal experience: "Forecaster A is excellent for Item X, especially in winter", "Forecaster B has performed very well over the last few days", "Forecaster C has hit no jackpots, but usually achieves limited successes", and so on. This study concerns the proper choice of such factors: what characteristics of a forecast should be considered, and how does this choice affect the accuracy of the combined forecast?

From the theoretical point of view, the combining techniques of single-value forecasts can be explained by an optimal estimation theory (Deutsch, 1965; Liebelt, 1967; Schweppe, 1973; Maybeck, 1979), which has been widely applied in science and industry. The optimal estimator was provided as a computational function that uses measurements to deduce a minimum error estimate of the state in a system. This is possible by utilizing knowledge of system and measurement dynamics, assumed statistics of system noises and measurement errors, and initial condition information (Gelb, 1974). Many estimation procedures such as least squares, weighted least square, minimum variance, and minimum variance unbiased estimation procedures can be employed for optimal

* Corresponding author. Tel.: +82 2 880 8916; fax: +82 2 887 8354.
E-mail addresses: dae_il_jeong@ete.inrs.ca (D.I. Jeong), yokim05@snu.ac.kr (Y.-O. Kim).
1 Tel.: +1 418 654 3751.

0022-1694/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.08.028

estimation (Liebelt, 1967). These procedures are theoretically identical to the combining forecast methods. More modern estimation techniques such as the Kalman filter in the time domain and the kriging in the space domain were also developed based on the optimal estimation theory (Gelb, 1974; Isaaks and Srivastava, 1989).

Forecast combining techniques have been extensively studied in the field of econometrics ever since the seminal article of Bates and Granger (1969). More than 200 works on the topic were reviewed and summarized by Clemen (1989). He concluded that forecasts can be substantially improved by combining multiple individual forecasts, and laid out the most effective techniques discovered so far.

The concept of combining forecasts has also taken root in climate modeling. In this context it is usually based on the work of Vislocky and Fritsch (1995) and Krishnamurti et al. (1999), who developed the 'multi-model ensemble' and 'multi-model superensemble' methods respectively. Vislocky and Fritsch (1995) and Fritsch et al. (2000) showed that the simple average of model outputs (i.e., a multi-model ensemble) is a more skillful predictor than the best individual forecast. In practice, however, multi-model forecasts are usually created by fitting a weighted average of the individual forecasts; this is called a multi-model super-ensemble. This method has been widely used to produce accurate long-range climate forecasts (Krishnamurti et al., 1999, 2000, 2003; Kharin and Zwiers, 2002; Kumar et al., 2003; Coelho et al., 2004). However, the simple average may be a choice when many scenarios exist in an ensemble because estimating many parameters in the weighted average may be impossible. Makridakis and Winkler (1983) also illustrated that the combined forecast decreases most of the possible errors but they did not obtain further significant gains when they employed more than five individual forecasts. Raftery et al. (2005) proposed a Bayesian Model Average (BMA) method for combining predicted distributions from different ensemble models. The BMA pdf (probability density function) forecast can be provided by a weighted average of forecasted pdfs based on each of the individual model forecasts. Vrugt et al. (2006) developed a multi-objective formulation for ensemble forecasts using BMA. They applied the approach for 48-h ensemble data of surface temperature and sea level pressure, and multi-model seasonal forecasts of temperature. Sloughter et al. (2007) also applied the BMA for probabilistic quantitative precipitation forecasting whose predicted pdfs are not approximated by normal distribution.

There have been a few applications made in the field of hydrologic forecasting as well. McLeod et al. (1987) adopted combining techniques to predict flow. They made several individual forecasts using a conceptual rainfall–runoff model and Box and Jenkins (1970) time series models, then combined these forecasts by a weighted average based on the error covariance. Although their experiment tested only a few different combining methods, they found significant improvements in the performance of the proposed combining method. More recently, Shamseldin et al. (1997) combined the outputs from five rainfall–runoff models and 11 catchments using three different approaches: a simple average, a weighted average, and a neural network. The combined daily discharge estimates were consistently more accurate than the best individual estimate, especially for the weighted average and the neural network. This work was extended by Shamseldin and Oconnor (1999) who proposed a real-time model capable of combining one conceptual and two empirical (black-box) models. They proved that updating continuously discharge forecasts by a real-time combining method could improve the discharge forecasts of individual rainfall–runoff models.

Coulibaly et al. (2005) used the weighted average method to improve their hydrologic forecasts, a technique which proved skillful up to 4 days ahead. The individual forecasts in this case were based on three quite different models: a nearest-neighbor model, a conceptual model, and an artificial neural network. Combining approaches also have been employed to probabilistic forecasting problems. Georgakakos et al. (2004) illustrated that the multimodel ensemble mean was in general an improvement over the best single-model simulation. Ajami et al. (2006) extended the work of Georgakakos et al. (2004) by testing several advanced multi-model combining techniques for hydrological forecasting. Our objective also is to provide more statistically sophisticated multimodel combining techniques although we deal with deterministic forecasts. Kim et al. (2006) tested five different combining methods for two rainfall–runoff model simulation outputs, with the goal of developing a more accurate ensemble streamflow prediction system. The combination methods considered in this work were the simple average, constant coefficient regression, switching regression, the sum of squared error, and the artificial neural network methods. These were applied to two individual models: a conceptual rainfall–runoff model called TANK, and a black-box rainfall–runoff model based on an ensemble of neural networks. Among these methods, the sum of squared error to obtain time-varying weights performed best with respect to the root-mean-square error. The first part of the present study will introduce the most common methods of combining hydrologic forecasts. Although Kim et al. (2006) also provided a broad review of those combining methods commonly used in economic forecasting and dealt with some hydrological applications, they could not provide an appropriate combining guideline for hydrologic forecasters. Recently, Marshall et al. (2007) introduced a hierarchical framework for combining rainfall–runoff model outputs. This framework can estimate time-varying weights (probabilities) of constituent models. Moreover, they used the combining weights to select an appropriate hydrologic model for a certain catchment condition (e.g. low flows). However, their study focused on applying the new approach for combining hydrologic simulation outputs rather than providing a comprehensive combining guideline. Several studies employed the BMA approach to hydrologic forecasting. Duan et al. (2007) applied the BMA to develop probabilistic hydrologic predictions from multiple predictions produced by hydrologic models. Neuman (2003) and Ye et al. (2004) suggested a maximum likelihood BMA method and asserted that their approach is computationally feasible and applicable to a wide range of real-world hydrologic problems when reliable prior information is insufficient.

As Armstrong (1989) criticized, there are few guidelines specific enough to be useful to forecasters because they did not successfully provide appropriate combining methods in real situations. Despite extensive studies of the topic of "combining", this remains true today. In economics, Menezes et al. (2000) provided a reasonable guideline for the use of combining forecasts. They focused on three properties of the forecast errors which were variance, skewness and serial correlation of forecast errors. The combination methods considered in their study, however, were only constant weighting methods based on a linear combination. Armstrong (2001) provided a more general guideline for economic forecasts. He provided eight key principles for combining forecasts and three conditions favoring combining forecasts. The key principles are to use different forecast models, forecasts from at least five methods, formal procedures for combining, equal weights when facing high uncertainty, weights based on evidence of prior accuracy, weights based on track records, if the evidence is strong, and weights based on good domain knowledge. The three conditions favoring combining forecasts are when there is uncertainty as to the selection of the most accurate forecasting method, uncertainty associated with the forecasting situation, and high cost for large errors. However, his guideline may be too general to be employed for hydrologists. The most appropriate combining method depends heavily on the