Advanced Engineering Informatics 33 (2017) 427–439

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning

Kristina Doycheva [a,*], Gordon Horn [a], Christian Koch [b], Andreas Schumann [a], Markus König [a]

[a] Ruhr-University Bochum, Germany
[b] University of Nottingham, United Kingdom

ARTICLE INFO

Article history:
Received 15 January 2016
Received in revised form 16 August 2016
Accepted 16 November 2016
Available online 24 November 2016

Keywords:
Precipitation forecast
Supervised machine learning
Pattern recognition
Uncertainty

ABSTRACT

Numerical weather forecasts, such as meteorological forecasts of precipitation, are inherently uncertain. These uncertainties depend on model physics as well as initial and boundary conditions. Since precipitation forecasts form the input into hydrological models, the uncertainties of the precipitation forecasts result in uncertainties of flood forecasts. In order to consider these uncertainties, ensemble prediction systems are applied. These systems consist of several members simulated by different models or using a single model under varying initial and boundary conditions. However, a too wide uncertainty range obtained as a result of taking into account members with poor prediction skills may lead to underestimation or exaggeration of the risk of hazardous events. Therefore, the uncertainty range of model-based flood forecasts derived from the meteorological ensembles has to be restricted.

In this paper, a methodology towards improving flood forecasts by weighting ensemble members according to their skills is presented. The skill of each ensemble member is evaluated by comparing the results of forecasts corresponding to this member with observed values in the past. Since numerous forecasts are required in order to reliably evaluate the skill, the evaluation procedure is time-consuming and tedious. Moreover, the evaluation is highly subjective, because an expert who performs it makes his decision based on his implicit knowledge.

Therefore, approaches for the automated evaluation of such forecasts are required. Here, we present a semi-automated approach for the assessment of precipitation forecast ensemble members. The approach is based on supervised machine learning and was tested on ensemble precipitation forecasts for the area of the Mulde river basin in Germany. Based on the evaluation results of the specific ensemble members, weights corresponding to their forecast skill were calculated. These weights were then successfully used to reduce the uncertainties within rainfall-runoff simulations and flood risk predictions.

1. Introduction

Flood forecasts in small and medium-sized river basins are usually based on precipitation forecasts. These forecasts are derived by deterministic numerical weather models. As a result, the flood forecasts depend on model physics, initial and boundary conditions. Since the meteorological and hydrological processes are very complex, these forecasts are associated with a high level of uncertainty.

Conventional precipitation forecasts lead to unsatisfactory results because of these uncertainties [29]. With the aim to limit the range of the uncertainties, ensemble forecasting methods and ensemble prediction systems (EPS) are being applied more and more often [28]. The ensemble forecasting process includes several simulation runs using one or multiple models with perturbed initial conditions, varied parameter sets, or different model physics [8]. The individual combinations of these models and conditions, which are used for the multiple simulations, are referred to as ensemble members. Hence, an ensemble forecast predicts a vector of values for one specific point in time at one and the same location. This vector of values should ideally represent the range of possible developments in the near future.

* Corresponding author.
E-mail addresses: kristina.doycheva@rub.de (K. Doycheva), gordon.horn@rub.de (G. Horn), christian.koch@nottingham.ac.uk (C. Koch), andreas.schumann@rub.de (A. Schumann), koenig@inf.bi.rub.de (M. König).

http://dx.doi.org/10.1016/j.aei.2016.11.001
1474-0346/

In general, it is assumed that the probability of an ensemble member leading to a reliable forecast is the same among all ensemble members and the diversity of the members compensates for errors, either due to lack of comprehensive initial data, accuracy and quality of the initial data, or due to model resolutions that do not sufficiently capture all features of the observed event [24]. However, in some cases the inclusion of less skillful members degrades the quality of the ensemble prediction. This could lead to underestimation or exaggeration of the risk of hazardous events, such as floods. For example, underrating the possibility of a flood due to considering too many "poor" ensemble members could result in a missing warning. Vice versa, overestimating the probability of a flood to occur would lead to many false alarms, which may encourage people to ignore warnings in the future [9].

In order to overcome these limitations, this paper suggest to improve the capability of ensemble prediction systems by introducing weights to the members based on their forecast skill. In meteorological context, the term forecast skill refers to the accuracy of a prediction to an observed event [1]. To apply a probabilistic assessment of meteorological ensemble members, their performance to describe an observed event in the past can be used. By post-processing such events, forecasters can assign smaller weights to members with a poor prediction quality (less skilled members) or completely exclude such members from the ensemble to improve the overall quality of the forecast.

Some statistical methods to determine the performance of the individual members have been presented. However, further research into improved methods is still needed. Recently, examples have been presented that demonstrate unexpected skill with common deterministic verification metrics, such as the Brier skill score or relative operating characteristics [16], For this reason, methods which are not purely based on statistical measures need to be investigated. Instead of statistical measures, the knowledge and experience of hydrology experts could be employed in order to evaluate the performance of different ensemble members. Usually, the assessment procedure is completely manual and time-consuming, because numerous forecasts including thousands of ensemble members need to be considered.

Yet, the limitations of the manual assessment procedure can be compensated by utilizing automated approaches for forecast assessment. However, the knowledge of these experts is a so-called tacit knowledge. Tacit knowledge is difficult to transfer to another person by means of writing it down or verbalizing it. As a result, the knowledge of the experts cannot easily or directly be transferred to a software system. Consequently, a kind of an automated training and learning system that maps the expert's knowledge onto a software system is required in order to accelerate the assessment and weighting procedure.

In this paper we present an approach to assess precipitation forecasts based on supervised machine learning and use higher quality forecasts to reduce the uncertainty of flood risk predictions. The paper is structured as follows. Section 2 presents existing relevant and related methods for ensemble member weighting. The methodology is presented in Section 3, while a case study is described in Section 4. In particular, the observed area, the ensemble prediction system used to forecast precipitation, and the hydrological runoff simulation model are presented. The paper concludes with a summary and an outlook on future research.

2. Related work

Several assessment approaches, mostly based on statistical measurements, have already been proposed. Garaud and Mallet [13] measured the quality of uncertainty estimation, and the reliability and the resolution of an ensemble using the Brier score, or scores derived from the rank histogram or the reliability diagram. Eckel and Walters [10] used the verification rank histogram to interpret and adjust an ensemble probabilistic quantitative precipitation forecasts. Krishnamurti [20] used a weighted multi model ensemble to produce one averaged forecast. The weights were obtained by using multiple regressions between the model forecast and the observed data during a training period. However, Hamill and Swinbank [17] addressed the need of reducing systematic errors in numerical weather predictions. According to their work, statistical post-processing methods may be applied for this purpose, but continuous research into improved methods is still needed.

Hamill and Juras [16] also presented examples that demonstrate unexpected skill with common deterministic verification metrics, such as the Brier skill score or relative operating characteristic. Raftery et al. [26] used Bayesian model averaging for statistical post processing of meteorological forecasts, which was later extended specifically for probabilistic quantitative precipitation forecasts by Sloughter et al. [30]. Beside Bayesian model averaging, Casanova and Ahrens [6] applied a simpler skill-based weighting method, using the normalized inverses of the mean-square errors of the forecast in a training period as weights. Gneiting et al. [14] introduced the ensemble model output statistic. This post-processing technique also uses multiple linear regression and can be applied to any ensemble system and forecast variable, including precipitation, with some modification. Weusthoff et al. [34] also tried to classify precipitation forecasts into good and bad members. Although they also used a spatial approach, they applied an optical flow technique [18] for their classification process.

Brochero et al. [5] applied backward greedy selection to assess the weight of each model within a subset of members. Williams et al. [35] proposed a technique based on a machine learning method that creates random forests to compare the utility of various predictors and identify a subset that may be used for storm prediction. Random forests were used to identify regimes representing different types of geographical locations. Gagne et al. [12] used machine learning algorithms to account for some of the spatial and temporal uncertainties in convective precipitation forecasts.

Multiple approaches to post-processing storm-scale ensemble precipitation forecasts were evaluated. Mallet et al. [22] applied machine learning algorithms for sequential aggregation of ozone forecasts. Based on past observations, weights for each model were produced by learning algorithms. The performance of individual models was measured with the root mean square error. Evans et al. [11] calculated the mean absolute error, the root mean square error, the spatial correlation and the fractional skill score for individual events to create subset ensembles, considering simulation performance. In Kou [19], a recognition method for detecting abnormal patterns in spatial weather data, such as extreme precipitation events, is presented. The local outliers were identified with the help of graph based algorithms and a spatial neighborhood analysis. To model the spatial relationships, k-neighborhood based similarity graph applications were employed. With the aim of mapping different spatial scales, but keeping the complexity manageable, these graphs were constructed iteratively by means of wavelet transforms. However, the main focus is on the detection of abnormal conditions and tracking them over time. Wealands et al. [33] presented methods available for the automatic comparison of spatial hydrological patterns, with the aim of evaluating hydrological models. The spatial similarity was identified and compared to that of predicted and observed patterns.

The difference between the methods presented above and our approach is that our method includes the tacit knowledge of the experts into the automated evaluation and assessment process of precipitation forecasts. Especially in discharge prediction it is