Effects of temporal resolution of input precipitation on the performance of hydrological forecasting

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.5194/adgeo-29-21-2011

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

**Advances in
Geosciences**

Effects of temporal resolution of input precipitation on the performance of hydrological forecasting

F. Wetterhall[1,2], Y. He[1,3], H. Cloke[1], and F. Pappenberger[4]

[1]Department of Geography, King's College London, London, UK
[2]Swedish Meteorological and Hydrological Institute, Norrköping, Sweden
[3]Tyndall Centre for Climate Change Research, University of East Anglia, Norwich, UK
[4]European Centre For Medium Range Weather Forecasts, Reading, UK

Abstract. Flood prediction systems rely on good quality precipitation input data and forecasts to drive hydrological models. Most precipitation data comes from daily stations with a good spatial coverage. However, some flood events occur on sub-daily time scales and flood prediction systems could benefit from using models calibrated on the same time scale. This study compares precipitation data aggregated from hourly stations (HP) and data disaggregated from daily stations (DP) with 6-hourly forecasts from ECMWF over the time period 1 October 2006–31 December 2009. The HP and DP data sets were then used to calibrate two hydrological models, LISFLOOD-RR and HBV, and the latter was used in a flood case study. The HP scored better than the DP when evaluated against the forecast for lead times up to 4 days. However, this was not translated in the same way to the hydrological modelling, where the models gave similar scores for simulated runoff with the two datasets. The flood forecasting study showed that both datasets gave similar hit rates whereas the HP data set gave much smaller false alarm rates (FAR). This indicates that using sub-daily precipitation in the calibration and initiation of hydrological models can improve flood forecasting.

1 Introduction

Numerical Weather Prediction models (NWP) produce operational forecasts that can drive hydrological models to produce flood forecasts. Uncertainties in NWPs come from for example difficulties in describing the initial state of the atmosphere and the chaotic behaviour of the system. These uncertainties can have huge consequences on the predicted weather development over time. Therefore, NWP model runs

have been complemented with Ensemble Prediction Systems (EPS) since 1992 (Palmer et al., 1994). The EPS are created by running a simpler version of the NWP with perturbations in the initial state of the atmosphere to create an ensemble of possible weather developments. EPS also gives a measure of the uncertainty in the NWP (Buizza, 2008). This allows the operational flood forecasters to make a probabilistic assessment of the flood risk, but it also poses challenges on how to interpret these uncertainties and how they are dealt with institutionally (Demeritt et al., 2010; Nobert et al., 2010; Cloke and Pappenberger, 2009).

Forecasts issued from EPS are typically issued once or twice a day at a 6-h time step. Ideally, the hydrological models to be used in connection with the 6-hourly meteorological forecasts should be calibrated with observed precipitation against discharge both at a 6-h interval. Due to lack of observed precipitation at a 6-h temporal resolution, calibration is often carried out for the hydrological models using daily or 12-h observations. Furthermore, observed precipitation is often collected with a daily or 12 h resolution, typically at 09:00 or 21:00 UTC, and this creates a discrepancy when the models are run operationally with 6-hourly forecast data, which are issued at time intervals starting at 12:00 UTC (Fig. 2). There are an increasing number of stations collecting data on a high temporal resolution (hourly or sub-hourly) but there are still too few to capture the spatial pattern of rainfall events. However, for small catchments hourly precipitation data can give important information on storm events that are potentially flood inducing which could otherwise be missed in the daily precipitation series.

Earlier studies have shown the benefit of using high temporal resolution station data (e.g. Li et al., 2010; Pluntke et al., 2010) or assimilating radar data to create data sets of hourly time resolution or higher for input to hydrological models (e.g. Biggs and Atkinson, 2010). The evaluation is usually done on an integrating factor, such as modelled runoff. Forecasts are usually evaluated against observational
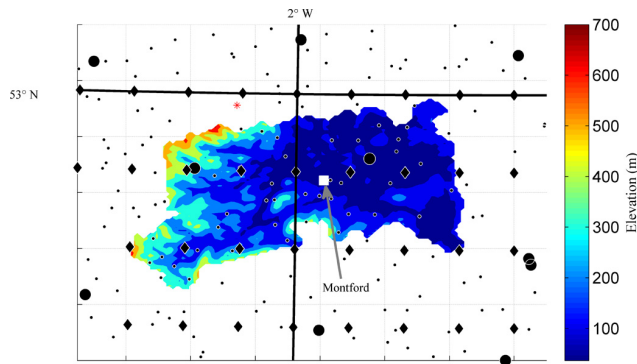
Fig. 1. Station data in the Upper Severn catchment showing daily stations (small black dots) and hourly stations (large black dots) used to interpolate the precipitation grid. The black diamonds are the location of the grid points from the ECMWF forecast and the white box denotes the location of the Montford station.

data to assess their quality and the benefit of using more than one NWP (He et al., 2009). However, in this study we take the opposite approach and evaluate the quality of the input precipitation using the forecast from the European Centre for Medium Range Weather forecasting (ECMWF) as benchmark.

This second part of the paper studies whether the use of hourly data in the calibration and spin-up of two hydrological models improves their ability to forecast floods. One of the models is then evaluated as a forecasting tool using HP and DP data sets for calibration and initialisation.

2 Data and methods

2.1 Precipitation data

Precipitation station data for the upper Severn catchment (ca. 4062 km^2) was obtained from the UK Meteorological Office MIDAS data base for the period 2006–2010 (UK Met Office, 2010). The time resolution was daily (09:00–09:00 UTC), 12-hourly (09:00–21:00 UTC, 21:00–09:00 UTC) and hourly. There are 40 stations reporting precipitation on 24 or 12 h and only 2 hourly stations located within the catchment (see Fig. 1). The forecast precipitation used was the 51 member ensemble (50 perturbed + 1 control run) from the ECMWF operational ensemble prediction system (EPS; Vitart et al., 2008). The forecast is issued at 12:00 UTC each day at a 6-h interval for the next 240 h. The spatial distance of the ECMWF grid point is approximately 17 km and 28 km in the Easting and Northing direction, respectively.

The precipitation station data was first disaggregated to a 1×1 km grid resolution using a simple linear regression method. This step was required as one of the hydrological models runs at 1×1 km grid resolution. In the case of the hourly station data, the neighbouring hourly station was used
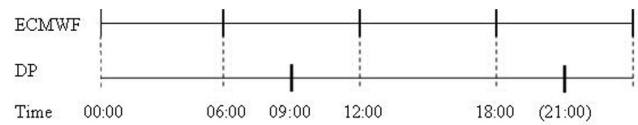
Fig. 2. Disaggregation of daily data to match the temporal resolution of the ECMWF data. The above line indicates when ECMWF forecasts are issued and the bottom line the time of day when 24- and 12-hour data is collected.

if it was closer than either of the two located inside the catchment. The hourly precipitation data within each corresponding 6-h forecast time interval was then aggregated to make up a 6-hourly time series. The 24- and 12-h data sets, on the contrary, were disaggregated to match the 6-hourly ECMWF EPS data by splitting each value into 6-h intervals. Where the time intervals overlapped, for example between 06:00 and 12:00, where the MIDAS data is reported at 09:00, the aggregation was made by averaging the MIDAS precipitation from the two immediately adjacent time steps (Fig. 2).

2.2 Evaluation of precipitation

The evaluation metric for the input precipitation series was the Continuous Ranked Probability Scores (CRPS; Hersbach, 2000). CRPS is a common tool to evaluate ensemble data:

$$\text{CRPS} = \frac{1}{N} \sum_{n=1}^{N} \int_{-\infty}^{\infty} [F(x) - H(x - x_0)]^2 \, dx \quad (1)$$

where N is the number of forecasts, $F(x)$ is the cumulative distribution function (c.d.f.) $F(x) = p(X \leq x)$ of the forecasted precipitation x, x_o the observed precipitation, and $H(x - x_0)$ is the Heaviside function, which has the value 0 when $x - x_0 < 0$ and 1 otherwise. In order to quantify the skill of the probability score, a skill score was calculated as

$$\text{SS}_{\text{CRPS}} = 1 - \frac{\text{CRPS}_{\text{FP}}}{\text{CRPS}_{\text{RP}}} \quad (2)$$

where CRPS_{FP} denotes the forecast score and CRPS_{RP} is the score of a reference forecast of the same predictand. A skill score of 1 denotes a perfect forecast, whereas 0 (or negative values) mean that the forecast performs equal to (or worse than) the reference. The reference forecast in this study was made up of precipitation series of 10 consecutive days of observed precipitation, starting 60 days up until 10 days before the forecast day. The lead time of 10 days was chosen to match the forecast lead time. The evaluation was done over each 1km grid point as well as over the entire catchment as a whole to investigate the effects of the higher spatial resolution in the DP data.

2.3 Hydrological modelling

The hydrological models used were LISFLOOD-RR (Van der Knijff et al., 2010) and HBV (Lindström et al., 1997) rainfall-runoff models. The LISFLOOD-RR is a distributed model run at 1×1 km distribution and the setup of the model is described in He et al. (2009). HBV model was run in a lumped version over the whole catchment as well as a lumped version upstream of Montford station (Fig. 1). The two model setups were selected to assess the effects of using a spatially distributed model (LISFLOOD) and a lumped model (HBV). The evaluation was performed on the modelled runoff over the period 2006–2009 using a number of scores, including Nash-Sutcliffe coefficient (NS). A parameter sensitivity study consisting of 200 000 runs with the HBV model and 10 000 with the LISFLOOD over the study period was first carried out and the parameter sets which had NS>0.7 and with a total difference in runoff not greater than +/−10% were selected as behavioural models. From these runs, four objective functions were calculated to assess the performance of the models. The objective functions were (1) flow-weighted NS, (2) mean squared error of annual peak flow error, (3) peak over threshold (POT) and (4) fuzzy weighting. All objective functions except 4 were assessed in comparison with a reference flow to calculate skill scores as in Eq. (2). The "fuzzy weights" in 4 were given as 1 if the simulated flow was between +/−10 % of the observed flow, and then a linear decline to 0 if simulated flow was between +/−30 %. The reference flow for the calibration period was the mean daily observed discharge.

2.4 Modelling flood events

The two model setups for the HBV model were then tested with the ensemble output from the ECMWF forecast model with 51 ensembles. The 10 best parameter sets from the calibration were selected using the POT and peak error scores as objective functions giving a total of 510 ensemble runs. The HBV model was evaluated at the Montford station, and in the forecast mode we used the Environment Agency of England and Wales (EA) warning level of 300 m³/s to identify a flood event. The evaluation of the flood forecasting was done through a contingency table using Hit Rates (HR) and False Alarm Rates (FAR) for each time step. HR is defined as the number of positive warnings (hits) divided by the number of missed events. FAR is defined as the number of false alarms (negative warnings) divided by the number of correct negatives (no warning).

3 Results

3.1 Evaluation of precipitation

The HP data set outperformed the DP data for lead times up to 4 days for the gridded data and 5 days for the lumped data

Table 1. Performance of the hydrological modes at Montford station. LF stands for the LISFLOOD-RR model and HP and DP are hourly and daily data, respectively. The best score for each rainfall-runoff model is highlighted in bold.

| Objective function	HBV-HP	HBV-DP	LF-HP	LF-DP
NS$_{flow}$	0.89	**0.91**	0.86	**0.88**
Peak error	0.92	**0.97**	**0.99**	0.97
POT	0.78	**0.81**	**0.69**	0.65
"Fuzzy flow"	0.24	0.24	0.32	**0.35**
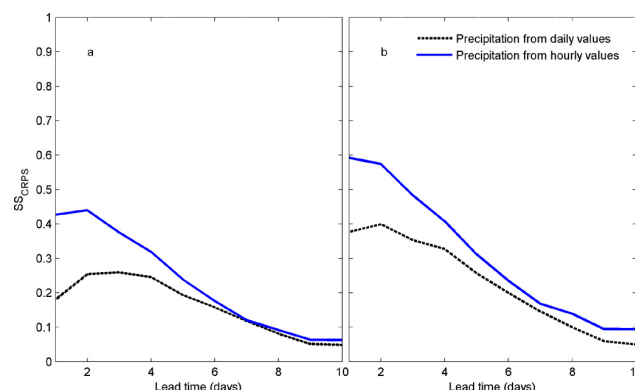
Fig. 3. SS$_{CRPS}$ for the entire period (1 October 2006–31 December 2009) for **(a)** 1-km gridded precipitation and **(b)** for the precipitation accumulated over the entire catchment. The black (dotted) line denotes DP and the blue (full) line HP.

when comparing SS$_{CRPS}$ (Fig. 3). Since the HP data has a much higher resolution in time, it is expected that the early lead times are better captured with hourly data than daily data, but for longer lead times the exact timing of precipitation event is not so important for the scores as the signal-to-noise ratio decreases. The analysis of the higher resolution (Fig. 3a) gives a lower overall score compared to when the precipitation first was averaged over the entire catchment (Fig. 3b). The better performance for the first 2 days of the HP data is consistent throughout the year, and for a majority of the months the HP outperforms DP up to 4 days (Fig. 4).

3.2 Hydrological modelling

The scores from the calibration of the hydrological models were similar for the different input precipitation data (Table 1). The scores should not be seen as a comparison between the different models, but rather how the different input precipitation affects models performance during calibration. The DP data gives the best performance for the HBV model, whereas the peaks are better modelled with HP using LISFLOOD and the DP gives better scores for the other scores. The flood forecast case study using the HBV model at Montford station indicated that both HP and DP gave similar HR, but the HP data had half as many false alarms (Fig. 5).
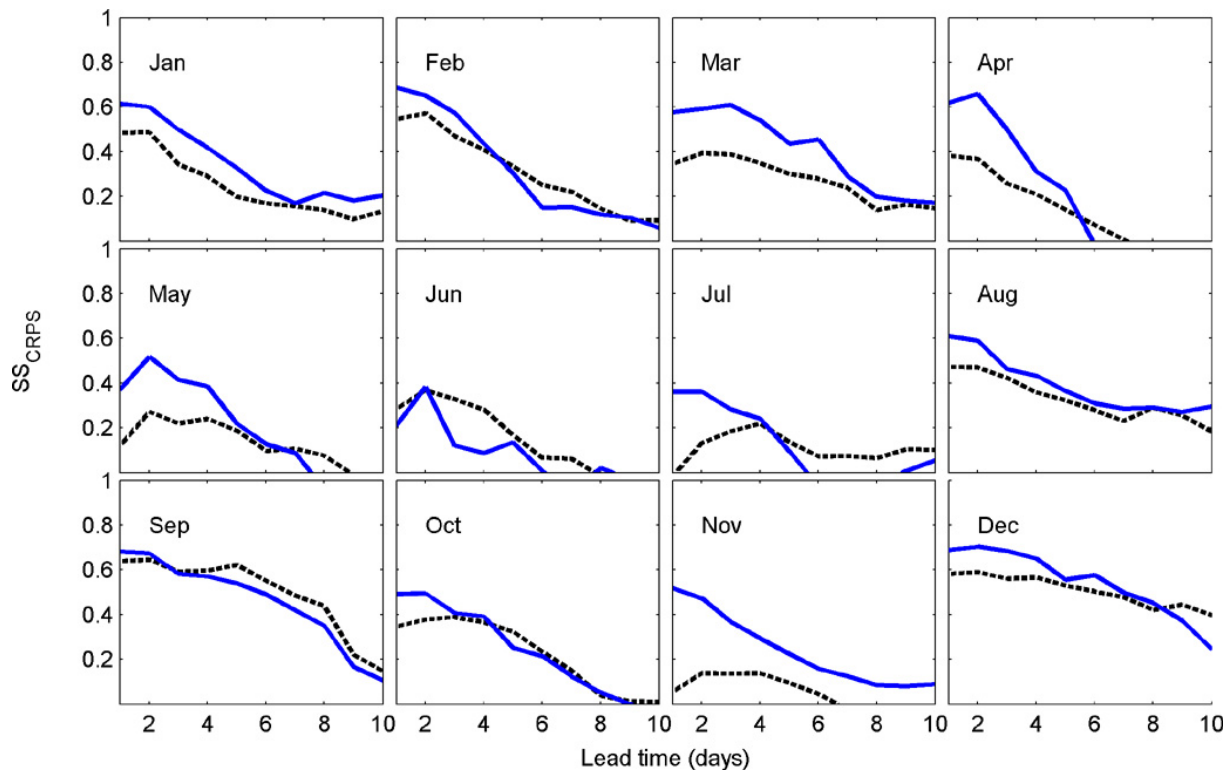
Fig. 4. Monthly S_{SCRPS} for the HP (blue) and DP (black dotted) precipitation series compared with to the ECMWF forecast for different lead times.

4 Discussion and conclusions

The precipitation calculated from HP showed higher skill than the data calculated from the DP, even though the spatial resolution is better represented by DP data (Fig. 3). The resolution of the ECMWF forecast is much coarser than the network of daily stations, so the potential advantage of a better spatial resolution in the DP data did not yield a higher score than the HD (Fig. 3a). The hydrological modeling results do not demonstrate benefit for either HP or DP (Table 1). It is difficult to state whether the effect of the better spatial resolution of DP is compensated by the better temporal resolution of HP. The results from the LISFLOOD model suggest that the higher temporal resolution of HD improves the peak discharge performance, where this is not evident in the HBV model. The fact that the HD dataset was based on very few stations (2 within the catchment and some adjacent stations) was not crucial for the model performance neither in calibration of the flood forecasting study implies that the timing of precipitation is very important in small catchments.

The results from the ECMWF flood forecast runs indicated that there is a benefit in calibrating the model using high temporal resolution data since this setup gave fewer false alarms (Fig. 5). The HR was similar, which means that the HP can be used to initiate the model without losing information on flood events. A low false alarm rate is very important for
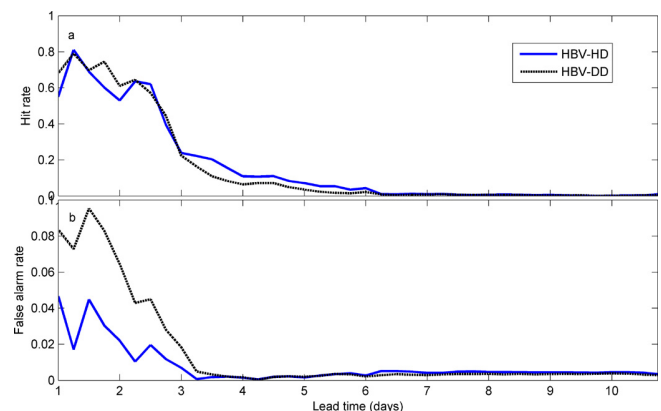
Fig. 5. Hit rate **(a)** and False alarm rate **(b)** as a function of lead time for HBV models calibrated and initiated by HP (blue lines) and DP (black dotted lines).

any flood forecaster since a false flood alert is costly. There is therefore a potential for better forecasts using the HP data for training and initialization of the flood forecast model. The best parameter values from the HP and DP calibration were very similar, so the improvement in FAR is probably because of better initial conditions for the hydrological model.

The main conclusions from this study are:

– forecasts can be used as an inverse tool to test the quality of different data sources

– precipitation calculated from hourly station data outperforms data calculated from daily stations when compared with the ECMWF forecast on a 6-h time step

– the performance of two hydrological models showed similar performance using daily and hourly precipitation during the calibration

– using hourly precipitation data decreased the false alarm compared with daily data rate without deteriorating the hit rate.

Even though the results from this study are promising there is still need for further studies to draw more general conclusions. The effects might be very local, and it is not certain that other catchments respond similarly. The number of hourly stations in this area was very scarce, so to draw more general conclusions either more catchments should be tested, or choose an area with a denser network of hourly stations. However, since HP data performed better against DP there is a potential benefit of using high resolution temporal precipitation in combination with EPS to achieve the best flood forecast. Also, assimilating other data sources, for example radar data, might lead to even better precipitation input maps. But to take full advantage of such improvements in space and time would require EPS forecast systems that produce precipitation fields of a better spatial and temporal resolution than is available today.

Edited by: J. Thielen-del Pozo
Reviewed by: two anonymous referees

References

Biggs, E. and Atkinson, P.: A comparison of gauge and radar precipitation data for simulating an extreme hydrological event in the Severn Uplands, UK, Hydrol. Process., doi:10.1002/hyp.7869, 2010.

Buizza, R.: The value of probabilistic prediction, Atmos. Sci. Lett., 9, 36–42, 2008.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375, 613–626, 2009.

Demeritt, D., Nobert, S., Cloke, H. L. and Pappenberger, F.: Challenges in communicating and using ensembles in operational flood forecasting, Meteorol. Appl., 17, 209–222, 2010.

He, Y., Wetterhall, F., Cloke, H. L., Pappenberger, F., Wilson, M., Freer, J. and McGregor, G.: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, Meteorol. Appl., 16, 91–101, 2009.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather Forecast., 15, 559–570, 2000.

Li, M., Shao, Q., and Renzullo, L.: Estimation and spatial interpolation of rainfall intensity distribution from the effective rate of precipitation, Stoch. Env. Res. Risk A., 24, 117–130, 2010.

Lindström G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, 1997.

Nobert, S., Demeritt, D., and Cloke, H. L.: Informing operational flood management with ensemble predictions: lessons from Sweden, Journal of Flood Risk Management, 3, 72–79, 2010.

Palmer, T. N., Buizza, R., Molteni, F., Chen, Y.-Q., and Corti, S.: Singular vectors and the predictability of weather and climate, Philos. T. R. Soc. A., 348, 459–475, 1994.

Pluntke, T., Jatho, N., Kurbjuhn, C., Dietrich, J., and Bernhofer, C.: Use of past precipitation data for regionalisation of hourly rainfall in the low mountain ranges of Saxony, Germany, Nat. Hazards Earth Syst. Sci., 10, 353–370, doi:10.5194/nhess-10-353-2010, 2010.

UK Meteorological Office: MIDAS Land Surface Stations data (1853–current), British Atmospheric Data Centre, http://badc.nerc.ac.uk/data/ukmo-midas, last access: 28 May 2010.

Vitart, F., Buizza, R., Alonso Balmaseda, M., Balsamo, G., Bidlot, J. R., Bonet, A., Fuentes, M., Hofstadler, A., Molteni, F., and Palmer, T. N.: The new VAREPS-monthly forecasting system: a first step towards seamless prediction, Q. J. Roy. Meteor. Soc., 134, 1789–1799, 2008.

Van der Knijff, J. M., Younis, J. and De Roo, A. P. J.: LISFLOOD: a GIS-based distributed model for river-basin scale water balance and flood simulation, Int. J. Geogr. Inf. Sci., 24(2), 189–212, 2010.