Acta Geophysica
https://doi.org/10.1007/s11600-020-00409-0

RESEARCH ARTICLE – HYDROLOGY

Impact of missing precipitation values on hydrological model output: a case study from the Eddleston Water catchment, Scotland

S. Ruman[1] · P. Krpec[1] · P. Rusnok[2] · A. R. Black[3] · M. Trizna[4] · T. Ball[5]

Received: 23 September 2019 / Accepted: 17 February 2020

Abstract
A hydrological model was applied to select the best infilling method of missing precipitation (1) and to assess the impact of the length of deleted and filled precipitation data (2). The model was calibrated and validated using the hourly observed discharges from two gauges located in the outlet of the catchment (62.34 km^2) and in the inner sub-catchment (2.05 km^2). Precipitation from four gauges was spatially interpolated over the overall catchment, while the sub-catchment used the precipitation from one gauge. Four scenarios of different lengths of deletion within three high-intensity events were established in the data of this gauge. Three infilling methods were applied and compared: substitution, linear regression and inverse distance weighting (IDW). Substitution showed the best results, followed by linear regression and IDW in both scales. Differences between methods were significant only in 8.3% and 19.4% of all cases (sub-catchment and catchment, respectively). The impact of length was assessed using the substitution only and by comparing differences in discharges and performance statistics caused by four scenarios. Higher differences in discharges were found on the catchment scale compared to the inner sub-catchment and were insignificant for all events and scenarios. The hypothesis that a longer length of deleted and filled data would lead to a greater error in discharges was wrong for 11.1% and 16.7% of all cases (sub-catchment and catchment, respectively). In several cases (33.4% sub-catchment, 27.1% catchment), the model produced better results using the time series with filled gaps compared to the configuration with observed data.

Keywords Missing values · High-intensity rainfall · Infilling methods · Hydrological model · Scotland

Introduction

Accurate precipitation data are essential for the understanding of hydrological processes, water resources planning, proposing flood protection or the mitigation of contamination (Beven 2012). Furthermore, precipitation data are the most important input for the hydrological models (Moulin et al. 2008). However, there remain high uncertainties when precipitation values in the time series are missing, often due to instrument or related failure (Wagner et al. 2012). Infilling methods provide a solution to fill in missing data; however, the right method has to be selected for each gauge due to its unique geographical location (Hwang et al. 2012). Estimations of missing time series are generally based on the measured data from gauges surrounding the targeted gauge (Cole and Moore 2008). The basic source of precipitation data remains the tipping-bucket gauge for the measurement of point rainfall depth (Cole and Moore 2008) often combined with the radar outputs (Jurczyk 2008; Pauthier et al. 2016; Boudevillain et al. 2016).

S. Ruman
stanislav.ruman@osu.cz

[1] Department of Physical Geography and Geoecology, Faculty of Sciences, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic

[2] Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, 30. Dubna 22, 701 03 Ostrava 1, Czech Republic

[3] Geography, School of Social Sciences, University of Dundee, Nethergate, Dundee DD1 4HN, Scotland, UK

[4] Department of Physical Geography and Geoecology, Faculty of Sciences, Comenius University in Bratislava, Mlynská dolina, Ilkovičova 6, 842 15 Bratislava 4, Slovakia

[5] Department of Archaeology, Anthropology and Geography, Faculty of Humanities and Social Sciences, University of Winchester, Sparkford Road, Winchester SO22 4NR, UK

Published online: 27 February 2020

Many studies propose different infilling methods for filling missing values of precipitation and consequently examine these methods using hydrological models (Heistermann and Kneis 2011; Lo Presti et al. 2010; Vicente-Serrano et al. 2010). Concerning the simple methods, substitution (Lo Presti et al. 2010; Vicente-Serrano et al. 2010), IDW (Dirks et al. 1998; Jurczyk 2008) and linear regression (Weisberg 2005) have all been employed for infilling missing values of precipitation. Vicente-Serrano et al. (2010) evaluated these methods during the homogenization of daily time series in north-east Spain and found the substitution as the best method, followed by the IDW and linear regression. Lo Presti et al. (2010) proposed filling gaps in daily precipitation time series in Italy by nonparametric regression in comparison with parametric regression and simple substitution. However, they found out the substitution method can be acceptable when the similarity value tends to be significantly high. Eischeid et al. (2000) used six different interpolation methods for completing daily time series in dependency of estimation biases for every gauge and every month in the western USA. They used, for the most part, multiple regression with the least absolute deviation criterion which outperformed IDW from remaining methods. Bárdossy and Pegram (2014) compared several methods for daily time steps in South Africa and proposed a new cupola-based method. Among the compared methods were the IDW, linear regression and substitution, while the IDW performed the best followed by linear regression and substitution. An artificial neural network (ANN) method with the regression tree was used in the study from the Appalachian Mountain, USA (Kim and Pachepsky 2010), where accuracy of the SWAT model was significantly improved using these infilling methods. Furthermore, the ANN approach was recently used along with the conventional cubic spline algorithm and multivariate linear regression method in the catchment of Southern England (Song et al. 2017) in high temporal resolution of rainfall rate estimation. As noted above, hydrological models were applied in this process, firstly to investigate the impact of missing precipitation data on the simulated outputs (dominantly discharges) and secondly to select the best infilling method (Singh 1997; Arnaud et al. 2002; Bárdossy and Das 2008; Moulin et al. 2008; Reusser et al. 2009; Hwang et al. 2012). Studies from France (Moulin et al. 2008) and Mexico (Arnaud et al. 2002) showed that by using the hydrological model, less detailed hourly rainfall input led to biased flow outputs. These biases can be compensated by model calibration applying the *effective parameters* approach (Beven 2006), but it results in higher output uncertainty. Although most of the studies analyse the effect of filling daily data series (Heistermann and Kneis 2011; Hwang et al. 2012; Kim and Pachepsky 2010), in higher temporal sub-daily resolution, there is a decrease in spatial correlation between gauges (Blenkinsop et al. 2016; Villarini et al. 2008; Lewis et al. 2018). Ficchi et al (2016) investigate the extent to which the performance of hydrological modelling is improved by short time-step data (6 min vs. daily rainfall) in the mesoscale French catchment. They reported significant improvement in performance with shorter time steps.

Generally, rainfall estimation errors increase with decreasing catchment size due to topographic variability, so cases of small catchments are the most problematic. This was shown in the study by Krajewski et al. (2003), where the time interval varied from 1 h to 5 min in the selected rain gauges of various environments. Furthermore, the small urban catchments also require smaller temporal resolution, which was the case of the Twenterand catchment in the Netherlands (Cecinati et al. 2017). We followed these studies and applied hydrological model MIKE SHE/MIKE11 running on an hourly time step to select the best infilling method for the three high-intensity rainfall events. The model was developed in the part of small-scale Eddleston Water catchment and its inner sub-catchment.

Analyses of the extent to which data infilling of precipitation input influences the outputs of hydrological models are rare. The study of Teegavarapu and Nayak (2017) examined the impact of filled precipitation datasets for different lengths of gaps. The result for the period from 1901 to 2006 at 53 rain gauges in south Florida indicated the data infilling does not introduce statistically significant bias in total annual precipitation values but may lead to underestimation of both magnitude and frequency of heavy and very heavy precipitation events. Furthermore, they reported the increase in bias with the increase in the amounts of missing data. We follow this study and assess the impact of gaps of various lengths during three high-intensity rainfall events on the hydrological model outputs.

Thus, the aims of this case study were as follows: firstly, to select the best (optimal) infilling method of high-intensity rainfall events using the hydrological model (1) and, secondly, to assess the impact of different lengths of infilled gaps in high-intensity rainfall events on the outputs of hydrological model (2). Both aims were solved on the sub-catchment and catchment scale to investigate the impact of the catchment area.

Study area

The Eddleston Water, near Peebles, Scottish Borders, UK, has a topography ranging from 150 to 700 m and the average rainfall of 980 mm year^{-1} (Fig. 1). The high-intensity rainfall events are primarily caused by frontal precipitation. The Kidston Mill stream gauge is located at the main river of the Eddleston Water (river kilometre 2.53), which is a right (16.39 km long) tributary of River Tweed. The average discharge is 1.27 m^3 s^{-1}, river slope is equal to 0.0069 m/m,

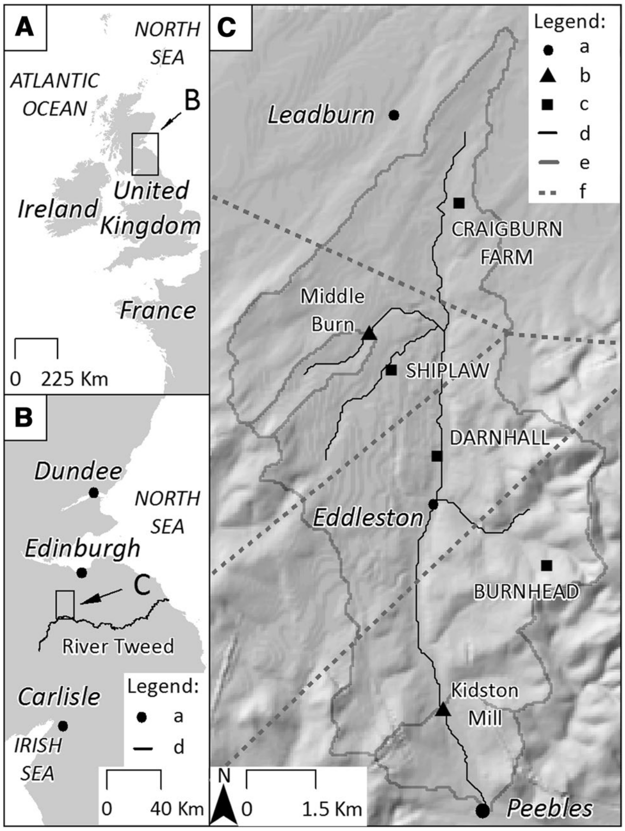

Fig. 1 Geographical position of the study area. A—localization within Western Europe, B—position within south-west Scotland, C—geographical situation of close surroundings of the Eddleston Water catchment. Legend: a—municipalities, b—stream gauges, c—rain gauges, d—waterways, e—sub-catchment borders, f—Thiessen polygons. Data sources: Eurostat, Ordnance Survey; Geographic Coordinate System: GCS OSGB 1936

and the sub-catchment area is 62.34 km^2 with the slope equal to 13.97%. The Middle Burn stream gauge (river kilometre 2.2) is located at the stream of the identical name with the length of 4.1 km, slope equal to 0.0215 m/m and average discharge equal to 0.06 $m^3 s^{-1}$. The stream is a right tributary of the Eddleston Water, located in the north-west part of the catchment (Fig. 1). The sub-catchment area is 2.05 km^2, and the sub-catchment slope is 7.59%. Geologically, it is mantled by complex post-glacial ice-margin deposits, surface strata (O'Dochartaigh et al. 2012) that overlay generally low-permeability Silurian greywackes (Pearce et al. 2014). The main river stem runs, from approximately its catchment midpoint, down a wide alluvium-infilled valley. Soils within the catchment are dominantly sandy loams (53%) followed by loamy sands (20%), peats (17%) and loam (10%; JHI 2014). Catchment land use is mostly grass (67%) followed by coniferous forest (10%, predominantly in the Middle Burn sub-catchment). Marsh, shrubs and mixed forest each account for 5% of the total catchment area. Urban development, stripes of arable land, water and broadleaf forest cover the remaining area. Since 2011, the catchment has been equipped with rain and stream gauges (Fig. 1). Research undertaken under the auspices of the Eddleston Water Project (Tweed Forum 2016) has investigated the effects of catchment management and restoration measures that have been implemented for both ecological improvements and as a means to alleviate local flood risk (Archer et al. 2013; Tweed Forum 2016). These measures were implemented at the end of August 2013 (Tweed Forum 2016) and thus did not influence the results of this study.

Methods

Hydrological model set-up, calibration and validation

A coupled rainfall–runoff/hydraulic model of the Eddleston Water catchment was developed using MIKE SHE/ MIKE 11 (DHI 2014). This established modelling system has been employed in a wide range of situations from small catchments or parts of catchments (e.g. Sahoo et al. 2006; Thompson et al. 2014; Thompson 2012) to large river basins (e.g. Andersen et al. 2001; Singh et al. 2011; Thompson et al. 2013, 2014). A 200 m $\times$ 200 m computational grid was employed, resulting in the catchment being discretized into 8000 cells. A digital terrain model (DTM) at 10 m grid resolution was created from contours with root mean square error/RMSE/$\pm$ 2.5 m. The model structure used the gravity flow method for the unsaturated zone formulation (MIKE SHE 2011). This comprised two layers for soils (JHI 2014) and bedrock (Hughes 1996) or superficial geology. A similar two-layer (upper zone: soils and lower zone: bedrock) approach was used in the finite difference saturated zone set-up with superficial geology represented as lenses. Hydraulic parameters for the unsaturated (saturated hydraulic conductivity, van Genuchten parameters) and saturated (horizontal and vertical hydraulic conductivity, specific storage and specific yield) zones were initially taken from the literature (Morris and Johnson 1967; O'Dochartaigh et al. 2012; MacDonald et al. 2012; Thompson 2012; Foster and Allen 2015) and were subject to manual calibration (Table 1).

The model used the finite difference approach for overland flow computation and the Kristensen–Jensen method for evapotranspiration. In the latter, a daily time series of reference evapotranspiration was computed (Allen et al. 1998) from the Darnhall meteorological station (as the climate data were available only in this station) and uniformly distributed over the catchment (see Fig. 1). Hourly precipitation from the four rain gauges for period 20/3/2011–23/6/2012 ('the study period') within the catchment was distributed using Thiessen polygons (Fig. 1). Although other methods

Table 1 Principal calibrated parameter values used in the model

Parameter	Average	Min	Max
Grid resolution (m)	200	–	–
Overland flow			
Manning *M*	3.0	1.0	20.0
Root depth (m)	–	0.0	1.5
Leaf area index	–	1.0	7.0
Crop coefficient	–	1.0	1.1
Detention storage (mm)	5.0	–	–
Unsaturated zone			
Saturated moisture content	0.4	0.4	0.5
Residual moisture content	0.0	0.0	0.1
Alpha	0.1	1.0E−02	0.4
N	1.6	1.5	1.8
SHC (m s^{-1})	1.8E−05	1.0E−13	7.0E−05
Saturated zone			
HHC (m s^{-1})	4.8E−06	1.9E−13	1.3E−05
VHC (m s^{-1})	9.7E−07	1.9E−13	3.7E−06
Specific yield	0.1	0.0	0.2
Specific storage	3.1E−02	1.0E−05	0.1

SHC saturated hydraulic conductivity, *HHC* horizontal hydraulic conductivity, *VHC* vertical hydraulic conductivity

of spatial interpolation of the precipitation were tried (e.g. Kriging), only the Thiessen polygon method allowed the overall area of the Middle Burn sub-catchment to gain precipitation from a single rain gauge. The MIKE 11 1D hydraulic model used the dynamic wave approximation of the St. Venant equations. Four branches were delineated using the Arc Hydro extension of ArcGIS (Maidment 2002) and DTM (Fig. 1). A total of 200 cross-sections were specified throughout the channel model and were based on the topographic survey (June 2013). Stream discharges were measured and rated at the Middle Burn and Kidston Mill gauge stations. The hydrological model of Middle Burn sub-catchment used data from a rain gauge situated directly in an adjoining sub-catchment area (the Shiplaw Burn), 0.9 km away. The second (Kidston Mill) integrates all four rain gauges as sources of rainfall. The maximum allowed time step for all components of the MIKE SHE model was set to one hour, while a fixed time step of 5 s was applied in the MIKE 11 hydraulic model. Due to the relatively short length of the simulation period and the limited availability of hydrological data to drive the model, period of 1.3 years (20/3/2011–23/6/2012) of rainfall and reference evapotranspiration inputs were repeated for a warm-up period immediately prior to the simulation period in order to establish initial conditions.

The calibration strategy was aimed to build the hydrological model able to reflect the hydrological response of the catchment not only during high-intensity events but also for the study period. The calibration procedure was as follows: the overall data for rainfall and discharge were subdivided into calibration (20/3/2011–19/3/2012) and validation (20/3/2012–23/6/2012) periods. The model was run and manually calibrated. The Nash–Sutcliffe efficiency index (NSE) was set as an objective function and was computed using the following equation:

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T}\left(Q_m^t - Q_o^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - Q_{\bar{o}}^t\right)^2} \tag{1}$$

where Q_m is simulated discharge, Q_o is observed discharge and $Q_{\bar{o}}$ is average observed discharge. Following Ritter and Muñoz-Carpena (2013), a threshold for satisfactory model performance, the value of NSE equal to 0.65 was deemed acceptable. This threshold was evaluated also for the selected events. During the calibration and validation, the *modified* Nash–Sutcliffe efficiency index (MNSE) and absolute total difference error were also assessed:

$$\mathrm{MNSE} = 1 - \frac{\sum_{t=1}^{T}\left(Q_m^t - Q_o^t\right)}{\sum_{t=1}^{T}\left(Q_o^t - Q_{\bar{o}}^t\right)} \tag{2}$$

$$\mathrm{ATD}\ \% = \left(\mathrm{abs}\left(\sum_{t=1}^{T} Q_o^t - \sum_{t=1}^{T} Q_m^t\right) / \sum_{t=1}^{T} Q_m^t \times 100\right) \tag{3}$$

The ATD % provides information about the volume changes and is crucial for flood volume balance. The MNSE uses the power of one, having a higher sensitivity to systematic errors (Krause et al. 2005). Furthermore, during every model run, water balance error (WBE) was calculated by the model as an additional measure of model performance.

Selection of the best gap-filling method and impact of the length of deleted and filled precipitation data

The Shiplaw rain gauge station was selected as a source of precipitation data for further steps. Three events of different lengths were chosen from the study period, using the criteria of maximal hourly rainfall, aiming for a spread of both duration and intensity. Based on this, three events with the maximal hourly rainfall over the study period were selected. The shortest was event 1 (63 h), followed by the lower-intensity multi-peak event 2 (366 h) and event 3 (72 h; Table 2). The event 1 produced the highest peak discharge equal to 20-year recurrence interval at Eddleston Village.

Four scenarios of data deletion within these events were established, to make the proportion deleted equally to 10%, 30%, 50% and 70% from the event total length. Deleted data were afterwards filled using the three methods of filling missing values of precipitation. In the simple substitution

Table 2 Observed characteristics of three events used in the study

Event	No. of hours	Kidston Mill				Middle Burn			
		Rainfall (mm h^{-1})		Discharge (m^3 s^{-1})		Rainfall (mm h^{-1})		Discharge (m^3 s^{-1})	
		Max	Cum	Max	Cum	Max	Cum	Max	Cum
1 (10/08/2011–11/08/2011)	63	6.6	56.7	16.6	500.3	6.6	60.2	1.1	28.8
2 (19/11/2011/–04/12/2011)	366	6.4	85.8	8.7	832.2	6.4	103	0.6	42.1
3 (02/01/2012/–05/01/2012/)	72	7.4	30.9	13.4	316	7	41.6	0.8	15.4

Max maximal hourly variable (rainfall/discharge), *Cum* cumulative value of a variable over the event

method, gaps were filled directly, using data obtained from the most similar station. Station similarity was assessed by correlation between stations expressed using Pearson's coefficient values (Lo Presti et al. 2010). For the linear regression, a substitute station was found using the same method as above. This station was then used as an explanatory variable for fitting a linear function by the least-squares method, and the equation obtained was then used for missing data prediction (Weisberg 2005). In the case of IDW, missing values were obtained as a weighted mean from all surrounding gauges, where the weight is proportional to the distance. The power value of three has been found most suitable for hourly data (Dirks et al. 1998), so this was adopted.

Four scenarios of synthetically deleted and filled precipitation time series of three events were varied in the hydrological model. While the start of the simulation for each of the events was the same as for the study period, the end of the simulation was set to the end of the particular event. The same model performance statistics (NSE, MNSE, and ATD%) were applied as for the model calibration and validation periods to select the best method of filling missing values of precipitation and to assess the impact of the event total length. However, three types of performance statistics were computed. The first type (T1) was calculated using Eqs. (1)–(3), applying the observed discharges and discharges produced by the hydrological model, which used four scenarios of synthetically deleted and filled precipitation data. Instead of using the observed discharges in Eqs. (1)–(3), the simulated discharges produced by the hydrological model using gap-free precipitation time series were applied for the second type (T2). The last type (T3) was calculated using Eqs. (1)–(3), but applying the observed and synthetically deleted and filled precipitation data instead of discharges. While the first two types (T1 and T2) allowed us to distinguish between the errors caused by inaccurate model and errors caused by each method of filling missing values of precipitation, the latter type was used to assess the transfer of the precipitation error to the model results. Performance statistics of all types were computed for the overall event length, not just for the deleted part. To select the best method of filling missing values of precipitation, all scenarios and events were evaluated together and median and interquartile ranges of performance statistics (Lo Presti et al. 2010) were compared. The impact of length was assessed for the best method of filling missing values of precipitation only. Mann–Whitney test was applied to the discharges to judge whether the differences caused by four scenarios of synthetically deleted and filled precipitation were significant.

Results

Hydrological model calibration and validation

While overall fits between the modelled and observed data were qualitatively good, simulated peaks were characteristically advanced compared with the observed flows (Fig. 2). These shifts in the timing of peak flows were higher in the Kidston Mill record; thus, better performance statistics were reported for the Middle Burn.

The NSE was equal to 0.84 (calibration) and 0.85 (validation) for the Middle Burn sub-catchment. Lower values were found for the Kidston Mill, when the NSE of all methods was equal to 0.74 and 0.73 for calibration and validation, respectively. Higher ATD% was found in Kidston Mill compared to Middle Burn, with the highest value for calibration equal to 22.8%. The best NSE values were reported for the event 3, followed by event 2 and event 1 in the Middle Burn, while the change in the second and third places occurred for the Kidston Mill. Detailed information for all performance statistics for the calibration and validation period and also for the three events is shown in Table 3. The average value for WBE was computed from all of the simulations. This was equal to 1.74% for the Middle Burn and 0.27% for the Kidston Mill catchment, indicating a lower computational error for the larger catchment.

Selection of the best filling method and evaluation of the impact of the length of deleted and filled precipitation data

The selection of the best method was based on the median and interquartile range of the NSE, MNSE and ATD% (Fig. 3). Substitution was found to be the best method, followed by linear regression and IDW for Middle Burn using the median as the criterion. All three types of performance

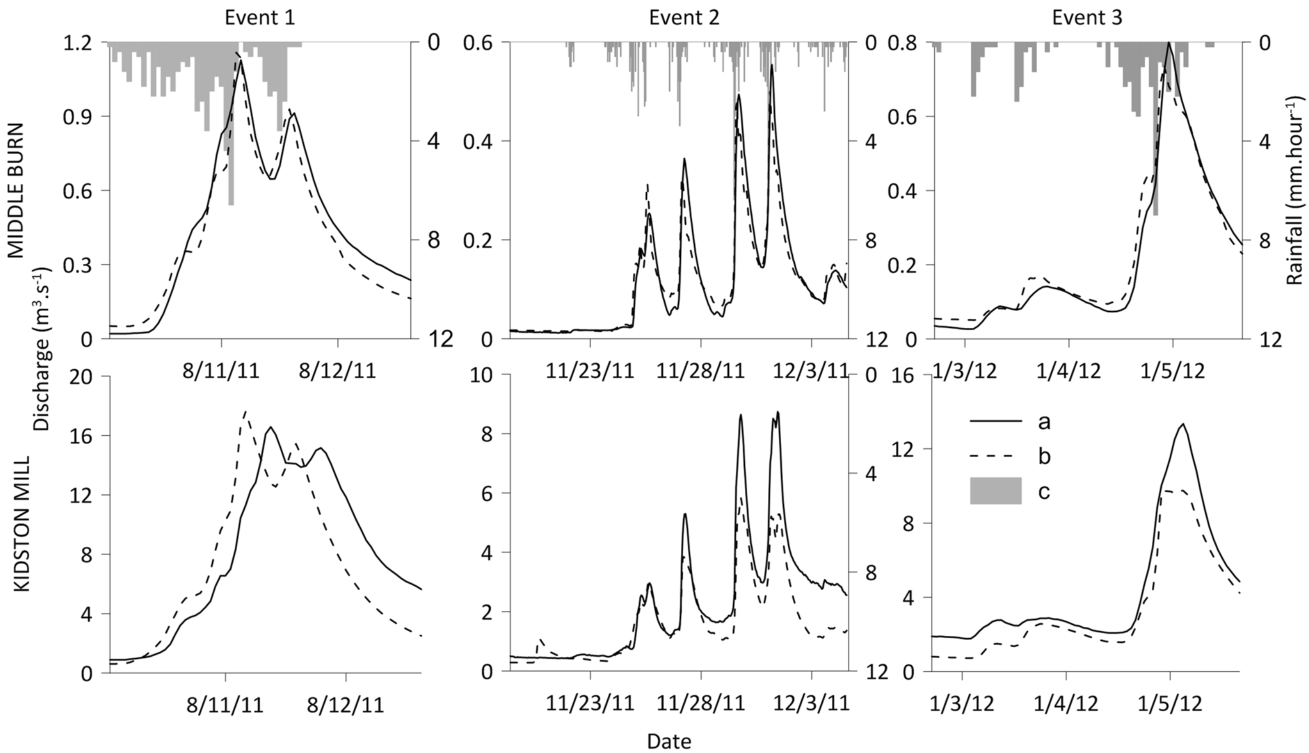

Fig. 2 Observed and simulated discharges in Middle Burn and Kidston Mill for three high-intensity events simulated by the hydrological model. Legend: a—observed, b—simulated, c—rainfall

Table 3 Hydrological model performance

Performance statistics	Period/event	Catchment	
		Middle Burn	Kidston Mill
NSE	Calibration	0.84	0.74
	Validation	0.85	0.73
	Event 1	0.94	0.65
	Event 2	0.93	0.74
	Event 3	0.97	0.86
MNSE	Calibration	0.65	0.49
	Validation	0.66	0.50
	Event 1	0.73	0.43
	Event 2	0.79	0.57
	Event 3	0.84	0.63
ATD%	Calibration	6.57	22.83
	Validation	6.72	12.18
	Event 1	7.09	9.27
	Event 2	0.82	24.49
	Event 3	5.93	23.54

statistics (T1, T2 and T3) agreed in this result. However, different results were found comparing the performance statistics T1 and T2 in the case of Kidston Mill. Applying the T1, the linear regression was favoured, while the T2 marked the substitution as the best method on the catchment scale.

Contradictory results among the best method selection were found applying the IQR criterion. For the Middle Burn, assessing the NSE, substitution produced the lowest IQR for all three types of performance statistics. Using IQR of MNSE and ATD%, the substitution was the best for the T2 and T3; however, IDW performed best for the T1 statistic. Similar contradictory outputs occurred on the catchment level: using the NSE and MNSE, the IDW performed the best applying the T1, while substitution was the best applying the T2. For the ATD%, the linear method produced the lowest IQR using the T1, but the substitution was the best applying the T2.

High Pearson's correlation coefficients (minimal value 0.95, maximal 0.99 from all events) were found when the three methods of filling missing values of precipitation were compared. The Mann–Whitney test showed the differences between methods were significant ($p < 0.005$) only in 8.3% of all of scenarios and events for the Middle Burn. These significant differences occurred during the longest event 2 between the IDW and linear regression for the scenarios 50% and 70% and between the IDW and substitution for the scenario 70%. Higher numbers of significant differences between the methods (19.4% of all cases) were found at Kidston Mill and occurred during the event 2, between the

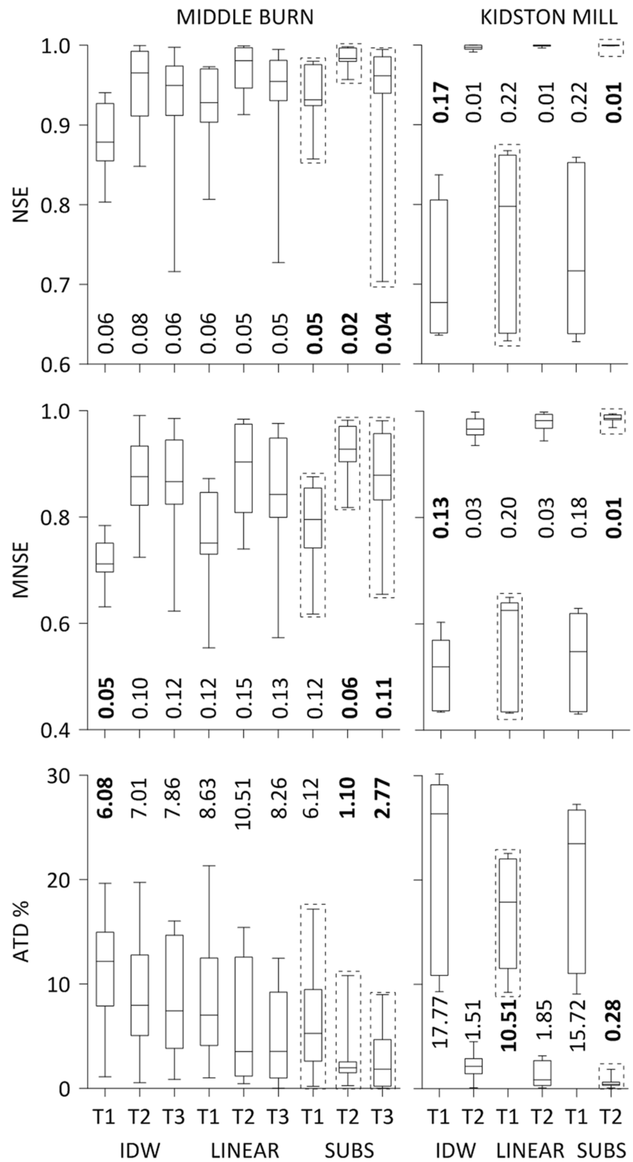

Fig. 3 Comparison of performance statistics of three methods infilling missing values of precipitation. Single box plot was created using the performance statistic of all events and all scenarios ($n=12$ for every single box plot; Whiskers = Max and Min). The best method based on the median is marked by the dotted rectangle. Numbers denote the values of IQR. Bolded values denote the lowest IQR among three methods

substitution and linear regression, IDW and linear regression for the scenarios 30, 50 and 70% and between the IDW and linear regression for the scenario 10%.

The transfer of precipitation error to the simulated discharges is visible in Fig. 3, by comparing three types of performance statistics (T1, T2 and T3). The performance statistics computed from precipitation data (T2) were generally better than statistics computed using observed and simulated discharges (T2), but worse than statistics computed using simulated discharges. Thus, the hydrological model lowered the error in the precipitation (T1 and T3).

The impact of the length of deleted and filled precipitation data was assessed using the substitution method only as this provided the best results. Differences between the simulation with observed precipitation data and four scenarios of deleted and filled precipitation data are visualized in Fig. 4. Higher differences in discharges were found on the catchment scale compared to the inner sub-catchment. Differences in performance statistics were higher on the sub-catchment scale and are shown in Fig. 5. These were caused by various scenarios of synthetically deleted and filled precipitation data for three high-intensity events and three infilling methods.

By assessing all events and performance statistics on the sub-catchment scale, in 33.4% of cases, the model produced better results when synthetically deleted and filled gaps were used in the hydrological model compared to the configuration with the observed precipitation data. Amount of cases was lower (27.1%) on the catchment scale. Most of these situations occurred for the event 3, followed by the event 2.

Furthermore, the higher length of deleted and filled data produced better results compared to the shorter length. This happened for 27.1% of all cases for the Middle Burn, while lower values (6.3%) were calculated for the Kidston Mill. This happened mainly during the event 3 on both scales. The most sensitive performance statistic for this detection was the ATD%.

Significant differences were reported for the event 2 and event 3 by the Mann–Whitney test for the Kidston Mill catchment, while all differences were marked as insignificant for the Middle Burn. This test used the simulated discharges produced by the model set-up with the four scenarios of deleted and filled precipitation data and the observed discharges.

Differences in performance statistics of the second type (T2) are shown in Fig. 6. Similarly, in results in Fig. 5, higher differences were reported for the Middle Burn than for the overall catchment and differences for the event 1 were the highest.

We further found out in several cases (11.1% of all cases for the Middle Burn) greater length of deleted and filled data led to better results compared to the shorter length. Higher values (16.7%) were calculated for the Kidston Mill. This happened dominantly during the event 3 on both scales. The most sensitive performance statistic for this detection was again the ATD%.

Differences were insignificant comparing discharges produced by four scenarios to the discharges simulated by the model with observed precipitation data for all events and scenarios.

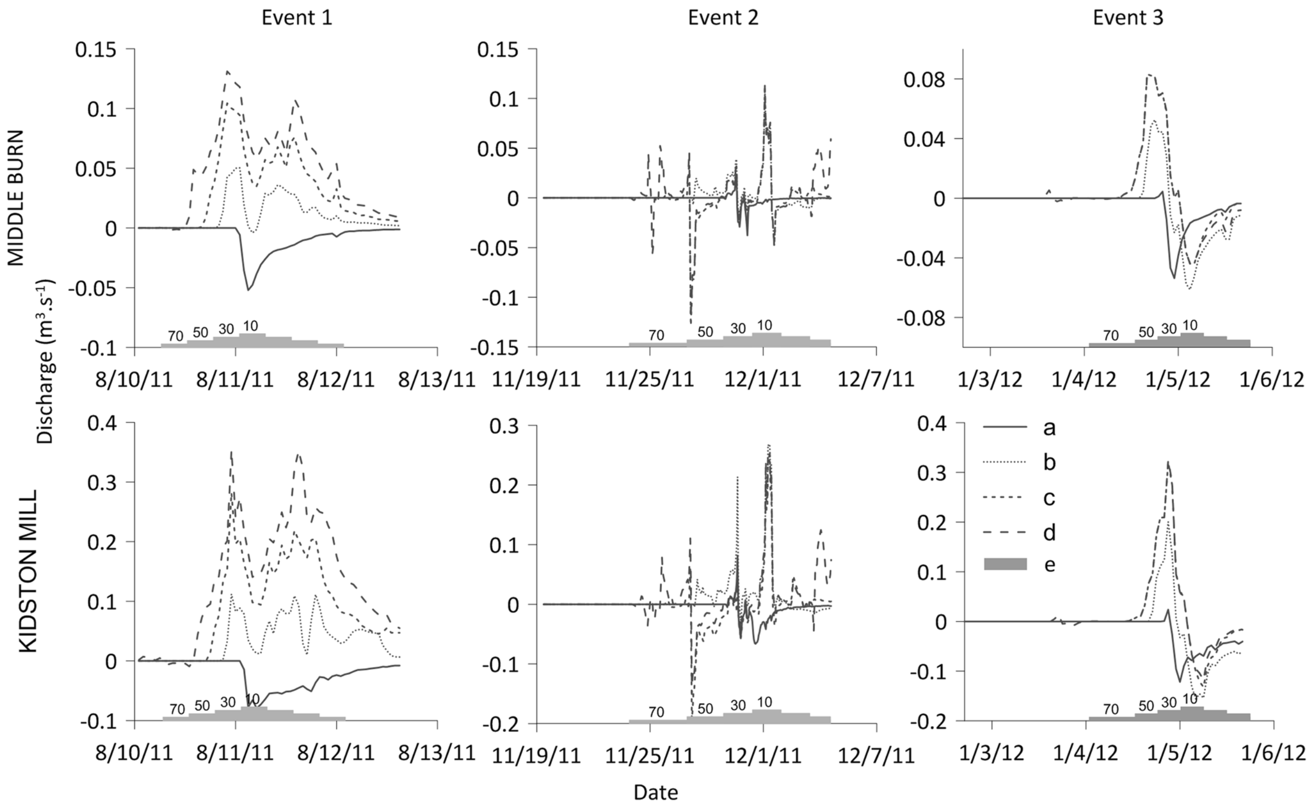

Fig. 4 Differences in discharges for four scenarios. Differences were computed between the simulation with observed precipitation data and four scenarios of deleted and filled precipitation data using the substitution. Legend: a—10% of deleted and filled data, b—30%, c—50%, d—70%, e—length of a particular scenario

Discussion

In this study, we applied the precipitation data to a fully distributed, physically based hydrological model. This type of modelling is based on the main premise; a high level of spatial resolution should lead to both improved representation of catchment behaviour and better simulation of the effects of changes in catchment processes and characteristics. As noted by Beven (2012), every single model requires its own model structure and own effective parameters (Vázquez and Hample 2014). Various models could produce different outputs as shown in the study of Huisman et al. (2009), where the effect of land cover change was examined; thus, results of this study (selection of the optimal infilling method) should be confirmed by different hydrological models.

Following the NSE criteria of model evaluation defined by Ritter and Muñoz-Carpena (2013), the model applied in this study was 'good' on the sub-catchment scale and 'acceptable' on the catchment scale for both the calibration and validation periods. All three events used in this study were marked as very good in the Middle Burn, while event 3 was good, and event 1 and event 2 were acceptable in the Kidston Mill. The computational error (WBE) for the Kidston Mill catchment were comparable to other studies (Foster and Allen 2015; Rahim et al. 2012). Higher errors in the WBE for Middle Burn could be caused by coarse-grid resolution (200 m), which might have been unable to account for sub-catchment hydrological processes. The length of the validation period was restricted to 3 months due to the hydrological data availability. Although a longer period (several years) would be essential to model variability of flood regime, we believe the results of this case study were not influenced by the length of the validation period as documented by fulfilling the criteria of model evaluation defined by Ritter and Muñoz-Carpena (2013). Various lengths of the IQR in the results could be caused by the model sensitivity of the input data reported by Beven (2006) and Vázquez and Hample (2014).

Three widely used methods for filling synthetically created gaps in the precipitation time series were assessed. Of the three methods used, the simple substitution produced the best results, followed by linear regression and IDW. This result is in agreement with the work of Lo Presti et al. (2010), where authors reported acceptable results by using the substitution when similarity among gauges was high. This can be a consequence of high correlation coefficient of the surrounding gauges. The correlation coefficient may be used as a weight in the IDW method rather than the

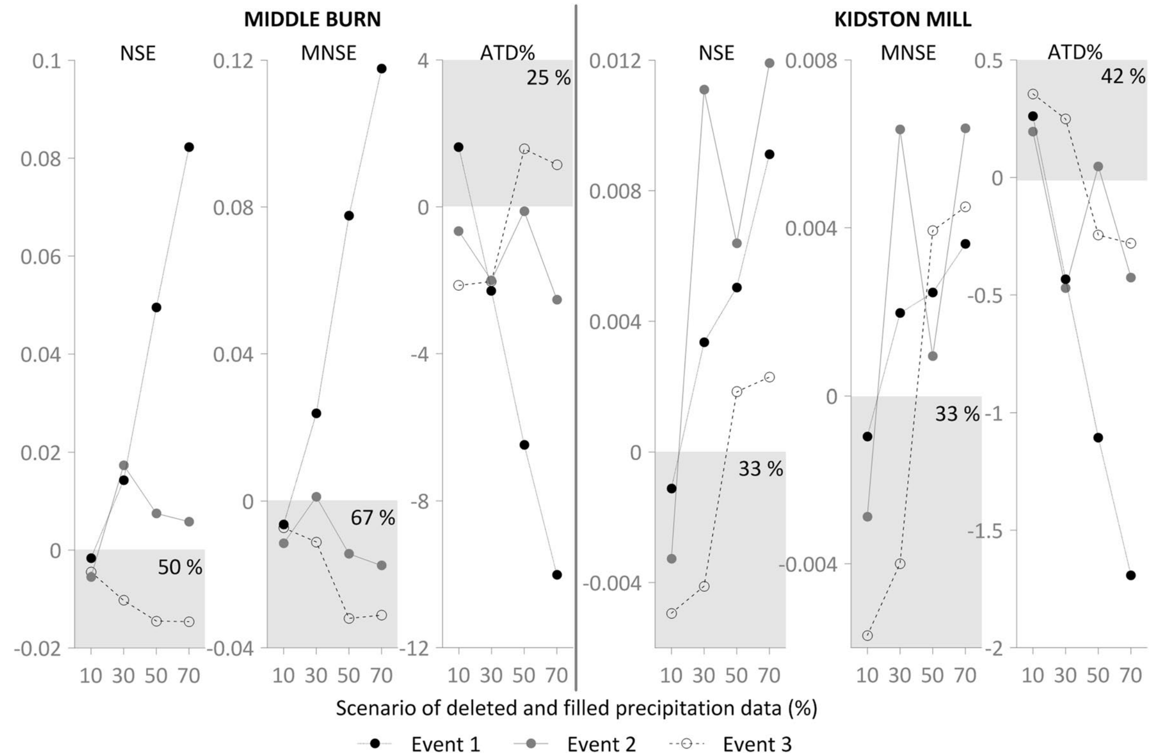

Fig. 5 Differences in performance statistics of type 1. Grey areas and percentage mark the simulations when the model produced better results using the time series with filled gaps compared to the set-up with gaps-free data

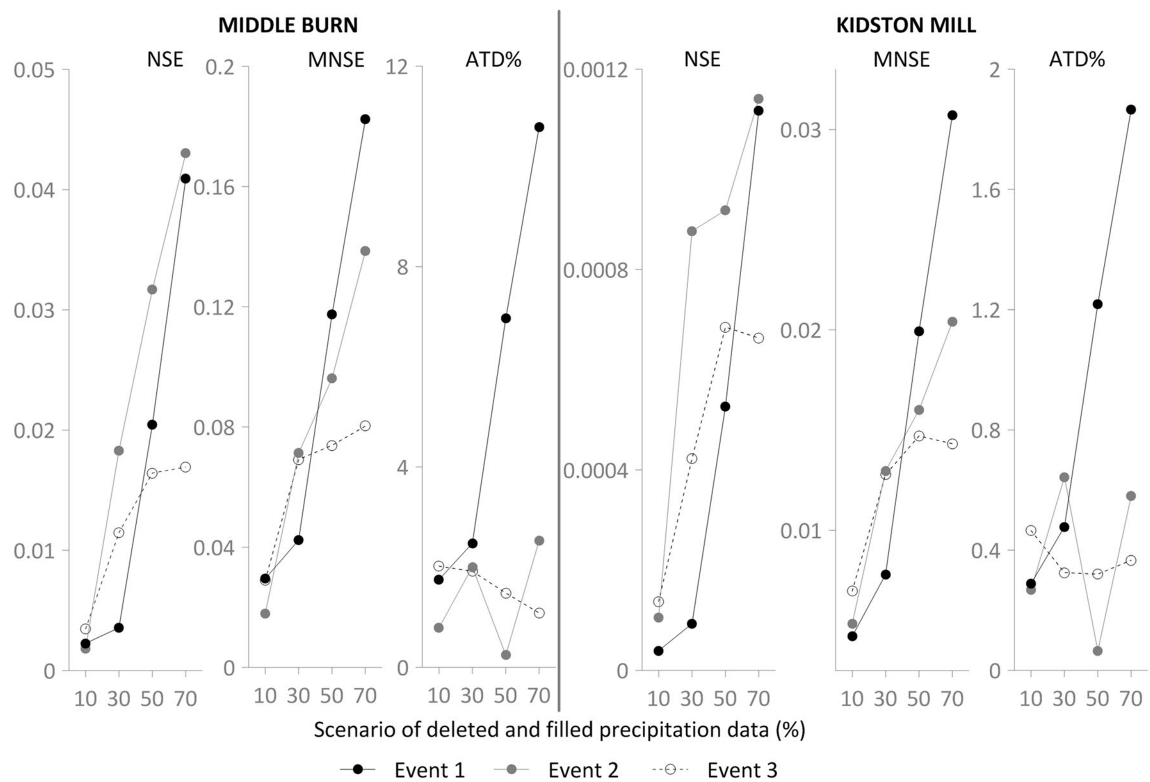

Fig. 6 Differences in performance statistics of type 2

distance (Vicente-Serrano et al. 2010) to improve results of this method.

Three performance statistics (NSE, MNSE and ATD%) were applied for both aims of this study. Contradictory conclusions were drawn in the selection of the best method and in assessing the impact of the length of deleted and filled precipitation data in the several cases using these statistics. This happened because each of the statistics was sensitive to different model (catchment) behaviour (e.g. high errors, water balance). This emphasizes the necessity of application of several performance statistics to obtain a holistic view of the outputs of the hydrological model (Moriasi et al. 2007; Ritter and Muñoz-Carpena 2013).

We computed three different types of performance statistics (T1, T2 and T3). While two of them were used for the hydrological model (T1 and T2), the latter was applied to the precipitation only. Transfer of precipitation errors to the results of the hydrological model was shown comparing all types of statistics. This is in agreement with other studies (Sun et al. 2000; Kuczera and Williams 1992; Kim and Pachepsky 2010). The model configured to use the synthetically deleted and filled precipitation data produced better performance than the model configured to use observed precipitation data, in several cases. In other words, the low-quality data lead to better model performance than high-quality input. This was revealed applying the T1, and it is a consequence of the effective parameters (Beven 2006) which were able to balance the inadequate representation of input data (Vázquez and Hample 2014; Alvarenga et al. 2016). Contradictory output in the selection of the best infilling method was found, comparing the T1 and T2 on the catchment scale. One remark needs to be done concerning the T1, T2 and T3. While T1 is independent of calibration results, T2 is based on the quality of the calibration (the better the calibration, the better the performance statistic). We propose to apply the T2 to select the best infilling method, as this type is using the simulated discharges produced by the model configured to use observed precipitation instead of the observed discharge.

Typically, distributed models employ the highest resolution of spatial data. However, the resolution of spatial data varies through the model (e.g. hydraulic properties of sediment and rocks, soils, land use data, digital terrain model, etc.) and is based purely on the data availability. The level of this availability allows the modeller to select more or less sophisticated model structure (Beven 2012). Furthermore, the application of the finest scale data does not necessarily provide the best agreement with observation. This was shown in the studies of Vázquez and Hample (2014) and Alvarenga et al. (2016), where datasets of lower quality (evapotranspiration in the first case and land cover in latter) produced superior model outputs over the higher quality data. In this study, the longer period of deleted and filled precipitation data used in the model leads to better model performance compared to shorter period in several cases, again because of effective parameters. Although it is highly possible, the more detailed calibration would eliminate this conflict; results of this study confirmed the problems associated with calibration of physically based, distributed hydrological models (Freeze and Harlan 1969; Beven 1993; Walker et al. 2003; Fatichi et al. 2016).

The length of deleted and filled precipitation data was assessed using four scenarios for three events of different lengths and rainfall intensities. In a majority of cases, we reported increase in bias with the increase in length of gaps, which is in agreement with the study of Teegavarapu and Nayak (2017). When comparing three infilling methods, significant differences were reported only for the longest event 2. Assessing the impact of the length, significant differences were reported for the longest event 2, followed by event 3. Thus, the length of the event had a crucial bearing to the modelled discharges. We further suppose the low magnitude of the events leads to the conclusion of this study: the impact of the length of deleted and filled precipitation data on the outputs of hydrological model is insignificant in a majority of cases. More events of higher recurrence interval should be applied to correctly investigate the impact of the length of deleted and filled precipitation data. This remark is based also on the study of Teegavarapu and Nayak (2017), where authors mark the events of heavy and very heavy rainfall as the most problematic.

The aims of the study were solved on two scales: overall catchment and inner sub-catchment. On the catchment scale, the differences between the simulated discharges produced by a model with a 'gaps-free' configuration and with four scenarios of deleted and filled precipitation data were higher than on the sub-catchment scale because of a larger area (Thiessen polygon), from where the information about the precipitation was spatially interpolated. Differences in performance statistics of both types (T1 and T2) were higher on the sub-catchment scale compared to the overall catchment. This is because the inner sub-catchment had the only one source of precipitation—the Shiplaw rain gauge station, from which data were synthetically deleted and filled. The overall catchment also used data from this station, but another three rain gauges with observed data were applied and spatially interpolated. Thus, the influence of missing precipitation data on the result of the hydrological model is greatest at the sub-catchment level and decreases with increasing catchment area due to synergic effects of other gauging stations. This finding is in agreement with the conclusion of Krajewski et al. (2003). However, more catchments should be examined in order to evaluate the impact of the length of deleted and filled precipitation data.

We applied a manual calibration strategy and approach of optimal parameter set to calibrate the model to hourly observed discharges in two gauging stations. Indeed, manual calibration remains a subjective approach, and a series of drawbacks in this approach were reported (Boyle et al. 2001; Vázquez and Hample 2014); the advantages of this approach are also known (Vaze et al. 2012). In the study of Vázquez and Hample (2014), authors found contrasting results for the manual and automatic calibration procedures.

Conclusion

We compared three infilling methods of precipitation for filling missing precipitation data. Our results showed the substitution provided the best results followed by linear regression and IDW, probably as a consequence of high correlation coefficient among rain gauges. Thus, in the case of gauges with the high correlation coefficient, the substitution can be used.

In our case study, the length of the event had a crucial bearing on the outputs of the hydrological model. However, only in a minority of cases, significant differences were reported between four scenarios of deleted and filled data, probably as a consequence of low magnitude of the events. Further analyses with events of higher magnitude should be carried out, and longer events should be evaluated to fully support this hypothesis.

Our results further indicate the data of lower quality (deleted and filled time series of precipitation) led to better model performance in several cases than the higher-quality data (original precipitation time series). This happened firstly when the hydrological model was fed by the original data and model performance was compared with the four scenarios and, secondly, when four scenarios of deleted and filled time series were compared between each other. Both cases were connected with the uncertainty associated with hydrological models, and modellers should be aware of this uncertainty and should carefully explain the results when the aim of the study is to compare the data of different qualities (not only the precipitation but also static catchment characteristics such as land use, soil texture and geological characteristics).

Lastly, the impact of the deleted and filled data on the model performance was higher on the sub-catchment scale. This is because the source of the precipitation data for this sub-catchment was from the gauge when the time series was deleted and filled. On the catchment scale, the impact was reduced by the synergic effect of four gauges. This emphasizes the necessity of close investigation of precipitation quality for the catchment where the source of data comes from the only single rain gauge.

Acknowledgements This work was supported by the research Project SGS01/UVAFM/2016 financed by the University of Ostrava and by Scottish Government Funding via Eddleston Water Project. We would like to thank Mr. Matěj Horáček for providing support to this work and to the Tweed Forum for the permission to use outputs of the gauged network set-up under the project.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration—guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56. Food and Agriculture Organization, Rome

Alvarenga LA, de Mello CR, Colombo A, Cuartas LA, Bowling LC (2016) Assessment of land cover change on the hydrology of a Brazilian headwater watershed using the Distributed Hydrology–Soil–Vegetation Model. Catena 143(2016):7–17

Andersen J, Refsgaard JC, Jensen KH (2001) Distributed hydrological modelling of the Senegal river basin—model construction and validation. J Hydrol 247(3–4):200–214

Archer NL, Bonell M, Coles N, MacDonald AM, Auton R (2013) Soil characteristics and landcover relationships on soil hydraulic conductivity at a hillslope scale: a view towards local flood management. J Hydrol 497:208–222

Arnaud P, Bouvier C, Cisneros L, Dominguez R (2002) Influence of rainfall spatial variability on flood prediction. J Hydrol 260(1–4):216–230

Bárdossy A, Das T (2008) Influence of rainfall observation network on model calibration and application. Hydrol Earth Syst Sci Discuss 3(6):3691–3726

Bárdossy A, Pegram G (2014) Infilling missing precipitation records—a comparison of a new copula-based method with other techniques. J Hydrol 519:1162–1170

Beven K (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. Adv Water Resour 16:41–51

Beven K (2006) A manifesto for the equifinality thesis. J Hydrol 320(1):18–36

Beven K (2012) Rainfall–runoff modelling: the primer, 2nd edn. Wiley, Chichester, p 455

Blenkinsop S, Lewis E, Chan SC, Fowler HJ (2016) Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. Int J Climatol 37:722–740

Boudevillain B, Delrieu G, Wijbrans A, Confoland A (2016) A high-resolution rainfall re-analysis based on radar–raingauge merging in the Cévennes-Vivarais region, France. J Hydrol 541:14–23

Boyle DP, Gupta HV, Sorooshian S, Koren V, Zhang Y, Smith M (2001) Toward improved streamflow forecasts: value of semidistributed modeling. Water Resour Res 37(11):2749–2759

Cecinati F, de Niet AC, Sawicka K, Rico-Ramirez MA (2017) Optimal temporal resolution of rainfall for urban applications and uncertainty propagation. Water (Switzerland). https://doi.org/10.3390/w9100762

Cole SJ, Moore RJ (2008) Hydrological modelling using raingauge and radar-based estimators of areal rainfall. J Hydrol 358(3–4):159–181

Dirks KN, Hay JE, Stow CD, Harris D (1998) High-resolution studies of rainfall on Norfolk Island. Part II: interpolation of rainfall data. J Hydrol 208:187–193

Eischeid JK, Pasteris PA, Diaz HF, Plantico MS, Lott NJ (2000) Creating a serially complete, national daily time series of temperature and precipitation for the western United States. J Appl Meteorol 39(9):1580–1591

Fatichi S, Vivoni ER, Ogden FL, Ivanov VY, Mirus B, Gochis D, Downer CHW, Camporese M, Davison JH, Ebel B, Jones N, Kim J, Mascaro G, Niswonger R, Restrepo P, Rigon R, Shen CH, Sulis M, Tarboton D (2016) An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J Hydrol 537:45–60

Ficchi A, Perrin CH, Andréassian V (2016) Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. J Hydrol 538:454–470

Foster SB, Allen DM (2015) Groundwater–surface water interactions in a mountain-to-coast watershed: effects of climate change and human stressors. Adv Meteorol 22:861805

Freeze RA, Harlan RL (1969) Blueprint for a physically-based, digitally-simulated hydrologic response model. J Hydrol 9:237–258

Heistermann M, Kneis D (2011) Benchmarking quantitative precipitation estimation by conceptual rainfall–runoff modeling. Water Resour Res 47(6):1–23

Hughes RA (1996) 1:63,360/1:50,000 geological map series. Sheet number: 24E. Sheet title: Peebles. 1: 50,000. I.G.S. Geological Survey of Scotland, Edinburgh

Huisman JA, Breuer L, Bormann H, Bronstert A, Croke BFW, Frede H-G, Graff T, Hubrechts L, Jakeman AJ, Kite G, Leavesley G, Lanini J, Lettenmaier DP, Lindstrom G, Seibert J, Sivapalan MG, Viney NR, Willems P (2009) Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) III: scenario analysis. Adv Water Resour 32:159–170

JHI (James Hutton Institute) (2014) Soil map (National Soil Map). 1:250,000. Vector data

Kim JW, Pachepsky YA (2010) Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. J Hydrol 394(3–4):305–314

Krajewski WF, Ciach GJ, Habib E (2003) An analysis of small-scale rainfall variability in different climatic regimes. Hydrol Sci J 48(2):151–162

Krause P, Boyle DP, Base F (2005) Comparison of different efficiency criteria for hydrological model assessment. Adv Geosci 5:89–97

Kuczera G, Williams BJ (1992) Effect of rainfall errors on accuracy of design flood estimates. Water Resour 28(4):1145–1153

Lewis E, Quinn N, Blenkinsop S, Fowler HJ, Freer J, Tanguy M, Hitt O, Coxon G, Bates P, Woods R (2018) A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr. J Hydrol 564:930–943

Lo Presti R, Barca E, Passarella G (2010) A methodology for treating missing data applied to daily rainfall data in the Candelaro River Basin (Italy). Environ Monit Assess 160(1–4):1–22

MacDonald AM, Maurice L, Dobbs MR, Reeves HJ, Auton CA (2012) Relating in situ hydraulic conductivity, particle size and relative density of superficial deposits in a heterogeneous catchment. J Hydrol 434–435:130–141

Maidment DR (2002) Arc Hydro: GIS for water resources. ESRI, Sacramento, p 203

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 50(3):885–900

Morris DA, Johnson AI (1967) Summary of hydrologic and physical properties of rock and soil materials as analyzed by the Hydrologic Laboratory of the U.S. Geological Survey. U.S. Geological Survey Water-Supply Paper 1839-D, p 42

Moulin L, Gaume E, Obled C (2008) Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. Hydrol Earth Syst Sci Discuss 5(4):2067–2110

O'Dochartaigh BE, MacDonald AM, Merritt JE, Auton CA, Archer N, Bonell M, Kuras O, Raines MG, Bonsor HC, Dobbs M (2012) Eddleston water floodplain project: data report. Nottingham, p 95

Pearce D, Rea BR, Bradwell T, McDougall D (2014) Glacial geomorphology of the Tweedsmuir Hills, Central Southern Uplands, Scotland. J Maps 10(3):457–465

Pauthier B, Bois B, Castel T, Thévenin D, Chateau Smith C, Richard Y (2016) Mesoscale and local scale evaluations of quantitative precipitation estimates by weather radar products during a heavy rainfall event. Adv Meteorol 2016:9

Ritter A, Muñoz-Carpena R (2013) Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. J Hydrol 480:33–45

Reusser DE, Blume T, Schaefli B, Zehe E (2009) Analysing the temporal dynamics of model performance for hydrological models. Hydrol Earth Syst Sci 13:999–1018

Sahoo GB, Ray C, De Carlo EH (2006) Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream. J Hydrol 327(1–2):94–109

Singh VP (1997) Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph. Hydrol Process 11(12):1649–1669

Singh CR, Thompson JR, Kingston DG, French JR (2011) Modelling water-level options for ecosystem services and assessment of climate change: Loktak Lake, northeast India. Hydrol Sci J 56(8):1518–1542

Song Y, Han D, Rico-Ramirez MA (2017) High temporal resolution rainfall rate estimation from rain gauge measurements. J Hydroinf 19:930–941

Sun X, Mein RG, Keenan TD, Elliott JF (2000) Flood estimation using radar and raingauge data. J Hydrol 239(1–4):4–18

Teegavarapu RSG, Nayak A (2017) Evaluation of long-term trends in extreme precipitation: implications of in-filled historical data use for analysis. J Hydrol 550:616–634

Thompson JR (2012) Modelling the impacts of climate change on upland catchments in southwest Scotland using MIKE SHE and the UKCP09 probabilistic projections. Hydrol Res 43(4):507

Thompson JR, Green AJ, Kingston DG, Gosling SN (2013) Assessment of uncertainty in river flow projections for the Mekong River using multiple GCMs and hydrological models. J Hydrol 486:1–30

Thompson JR, Green AJ, Kingston DG (2014) Potential evapotranspiration-related uncertainty in climate change impacts on river flow: An assessment for the Mekong River basin. J Hydrol 510:259–279

Tweed Forum (2016) The Eddleston Water Project Report (online). www.tweedforum.org. Accessed 20 Oct 19

Vaze J, Jordan P, Beecham R, Frost A, Summerell G (2012) Towards best practice model application. eWater Cooperative Research Centre, Australia, Bruce, p 46.

Vázquez RF, Hample H (2014) Prediction limits of a catchment hydrological model using different estimates of ET. J Hydrol 513:216–228

Vicente-Serrano SM, Beguería S, López-Moreno JI, García-Vera MA, Stepanek P (2010) A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. Int J Climatol 30:1146–1163

Villarini G, Mandapaka PV, Krajewski WF, Moore RJ (2008) Rainfall and sampling uncertainties: a rain gauge perspective. J Geophys Res 113:D11102

Walker WE, Harremoes P, Rotmans J, Van der Sluijs JP, Van Asselt MBA, Janssen P, Krayer von Krauss MP (2003) Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. Integr Assess 4(1):5–17

Weisberg S (2005) Applied linear regression, 3rd edn. Wiley, Minneapolis, p 336