



Originally published as:

Nied, M., Hundecha, Y., Merz, B. (2013): Flood-initiating catchment conditions: a spatio-temporal analysis of large-scale soil moisture patterns in the Elbe River basin. - *Hydrology and Earth System Sciences*, 17, 4, 1401-1414

DOI: [10.5194/hess-17-1401-2013](https://doi.org/10.5194/hess-17-1401-2013)



Flood-initiating catchment conditions: a spatio-temporal analysis of large-scale soil moisture patterns in the Elbe River basin

M. Nied, Y. Hundecha, and B. Merz

Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany

Correspondence to: M. Nied (manuela.nied@gfz-potsdam.de)

Received: 22 August 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 5 September 2012

Revised: 14 February 2013 – Accepted: 15 March 2013 – Published: 12 April 2013

Abstract. Floods are the result of a complex interaction between meteorological event characteristics and pre-event catchment conditions. While the large-scale meteorological conditions have been classified and successfully linked to floods, this is lacking for the large-scale pre-event catchment conditions. Therefore, we propose classifying soil moisture as a key variable of pre-event catchment conditions and investigating the link between soil moisture patterns and flood occurrence in the Elbe River basin. Soil moisture is simulated using a semi-distributed conceptual rainfall-runoff model over the period 1951–2003. Principal component analysis (PCA) and cluster analysis are applied successively to identify days of similar soil moisture patterns. The results show that PCA considerably reduced the dimensionality of the soil moisture data. The first principal component (PC) explains 75.71 % of the soil moisture variability and represents the large-scale seasonal wetting and drying. The successive PCs express spatially heterogeneous catchment processes. By clustering the leading PCs, we identify large-scale soil moisture patterns which frequently occur before the onset of floods. In winter, floods are initiated by overall high soil moisture content, whereas in summer the flood-initiating soil moisture patterns are diverse and less stable in time.

et al., 2010; Marchi et al., 2010). In order to capture the variety of large-scale flood generation mechanisms, flood events have been classified and analyzed according to their hydro-meteorological conditions along with their interactions between catchment state and meteorological conditions (e.g. Alila and Mtiraoui, 2002; Apipattanavis et al., 2010; Merz and Blöschl, 2003). These interactions vary from decade to decade (Alila and Mtiraoui, 2002), seasonally (Sivapalan et al., 2005; Merz and Blöschl, 2003; Parajka et al., 2010), and from event to event as well as from catchment to catchment (Merz and Blöschl, 2003). In addition to the occurrence and interaction of the hydro-meteorological conditions, their spatial patterns related to flooding need to be taken into account (Merz and Blöschl, 2003), which can be especially important in larger catchments (Merz and Blöschl, 2008).

On the regional scale, the automated classification of meteorological conditions has already identified a close relationship between the occurrence and persistence of meteorological circulation pattern types and floods (e.g. Bárdossy and Filiz, 2005; Jacobeit et al., 2006; Petrow et al., 2009; Prudhomme and Genevier, 2011; Parajka et al., 2010).

As far as hydrological catchment conditions are concerned, several studies identified soil moisture pattern types on the local or regional scale applying an automated classification (Kim and Barros, 2002; Jawson and Niemann, 2007; Korres et al., 2010; Ibrahim and Huggins, 2011; Perry and Niemann, 2007; Wittrock and Ripley, 1999). However, the analyzed soil moisture data (remotely sensed or ground-based point measurements) are either limited in their spatial extent covering a small ($< 1 \text{ km}^2$) study area (e.g. Perry and Niemann, 2007) and/or in their temporal resolution (monthly/annual values or a small number of subsequent days) (e.g. Jawson and Niemann, 2007; Wittrock and Ripley,

1 Introduction

Flood generation and magnitude are the result of a complex interaction between meteorological conditions, such as the amount and spatial distribution of precipitation or the inflow of warm air masses, and pre-event hydrological catchment conditions, such as soil saturation and snow water equivalent (Merz and Blöschl, 2008, 2009; Brocca et al., 2008; Parajka

1999). No studies are available that attempt to automatically classify the patterns of regional hydrological catchment conditions and to link them to flood initiation.

Complementary to the classification of meteorological conditions, we therefore propose the classification of the hydrological catchment conditions at the regional (Elbe) scale to get a probabilistic insight into the link between flood initiation and the hydrological catchment conditions. As soil moisture is a key variable of hydrological catchment conditions, we examine whether flood initiation in the Elbe River basin can be linked to specific soil moisture pattern types.

For the estimation of the hydrological catchment conditions concerning soil moisture, ground-based soil moisture measurements (e.g. time domain reflectometry, frequency domain reflectometry, gravimetric), remotely sensed soil moisture measurements (Brocca et al., 2009), model-based soil moisture (Norbiato et al., 2009; Merz and Blöschl, 2003) and surrogate measures such as mean annual precipitation (Merz and Blöschl, 2009; Merz et al., 2006), antecedent precipitation index (Merz et al., 2006; Brocca et al., 2009), Gradex method (Merz and Blöschl, 2008), and event runoff coefficient (Merz et al., 2006; Sivapalan et al., 2005; Merz and Blöschl, 2003, 2009) have been applied. In the present paper, the link between the hydrological catchment conditions and flood initiation in the Elbe Basin is investigated by using daily profile soil moisture simulated by a rainfall-runoff model and validated against remotely sensed soil moisture. It is assumed that the implemented rainfall-runoff model incorporates the hydrological processes that enable the estimation of soil moisture. Afterwards, a principal component analysis (PCA) and a subsequent clustering of the leading principal components (PCs) yield different pattern types. PCA is by far the most commonly applied method among the automated techniques to classify the structure of spatially variable data, and has also been applied in soil moisture pattern studies (Kim and Barros, 2002; Jawson and Niemann, 2007; Korres et al., 2010; Ibrahim and Huggins, 2011; Perry and Niemann, 2007; Wittrock and Ripley, 1999). In parallel, regional flood events are identified and linked to the derived pattern types.

The remainder of this paper is organized as follows: first, the study area and input data are described in Sect. 2. The methods to identify distinct types of daily soil moisture patterns and flood events are provided in Sect. 3. Section 4 describes the retrieved soil moisture pattern types, their characteristics as well as their relation to flood initiation. These results are discussed in the subsequent section. Section 6 concludes our findings and suggestions for future research.

2 Study area and data

2.1 Study area

The Elbe/Labe River (Fig. 1) originates in the Giant Mountains 1386 m a.s.l. in the Czech Republic, crosses northeastern Germany and reaches the North Sea after 1094 km. The Czech Republic and Germany are the main riparian states of the 148 268 km² large drainage basin. Negligible parts belong to Austria and Poland. About 50 % of the Elbe drainage basin has an elevation below 200 m a.s.l. One-third is hilly country with an elevation between 200 and 500 m a.s.l. The low mountain range (500–750 m a.s.l.) accounts for 15 % and the mountain range for less than 2 %. Major tributaries are the Moldau/Vltava contributing an average of 154 m³ s⁻¹ of river discharge (60 %) at its confluence with the Elbe River, the Eger/Ohře (38 m³ s⁻¹), the Mulde (67 m³ s⁻¹), the Saale (117 m³ s⁻¹), the Schwarze Elster (21 m³ s⁻¹) and the Havel (114 m³ s⁻¹). The Elbe River basin is situated in a transition zone between temperate (lower Elbe) and continental climate (middle and upper Elbe). Especially in the upper Elbe, the climate is strongly modified by the relief (IKSE, 2005). Mean annual precipitation in the river basin is 715 mm (1961–1990). However, there is a large variation within the basin. In the mountainous areas mean annual precipitation is above 1000 mm, whereas in the middle Elbe mean annual precipitation is around 450 mm. In the wintertime, precipitation falls as snow in the mountainous areas. Depending on snow depth and elevation, snow melts predominantly in March although it can persist until May, resulting in a snowmelt-influenced discharge regime. Mean annual evapotranspiration in the Elbe River basin is 455 mm (IKSE, 2005). In the highlands, thin cambisols are the main soil type, whereas in the lowlands, sandy soils and glacial sediments dominate. In the valleys, loamy soils are found. The western Elbe is covered by loess (chernozems and luvisols) (Hattermann et al., 2005). Land use is dominated by cropland (50.8 %), forest (evergreen 21.8 %, mixed 5.3 %, deciduous 3.1 %) and grassland (10.2 %). Settlements account for 6.5 % of the total basin area (CORINE European Environment Agency, 2000). Dams have been built in the Elbe headwaters and dikes have been installed along the river for flood protection purpose. The Havel region is strongly influenced by past mining activities. Previously observed flooding was predominantly generated by snowmelt in combination with rainfall in winter and spring in the upper Elbe. In summer, large-scale flooding due to long-lasting rainfall as well as small-scale flooding due to convective events were observed (IKSE, 2005).

2.2 Data

The data used in this study include climatic data as well as soil and land use information for driving a hydrologic model. Discharge observations are utilized for hydrologic model calibration and validation as well as flood event identification.

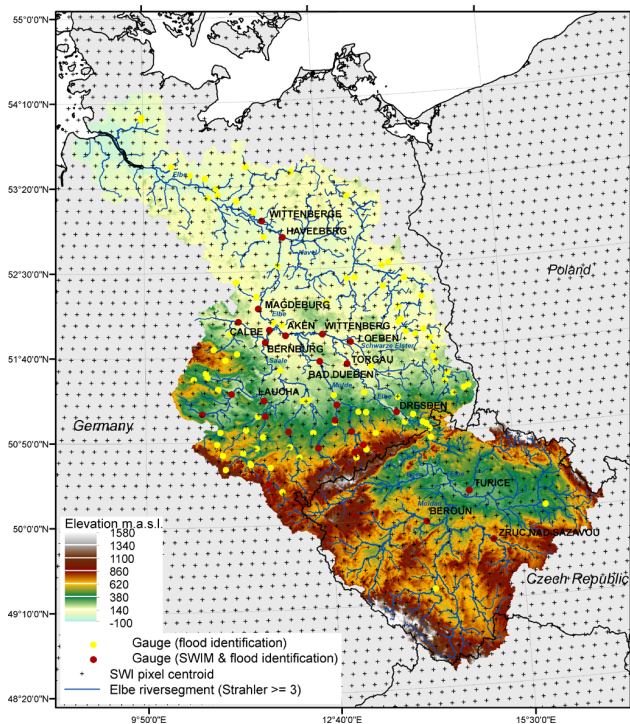


Fig. 1. Topographic map of the Elbe River basin. Yellow dots: discharge gauges applied in flood identification. Red dots: discharge gauges applied in flood identification as well as hydrologic model calibration and validation. Crosses: location of pixel centroids of the scatterometer data.

Scatterometer soil moisture estimates are used to validate the simulated soil moisture. The analysis time period is 1 November 1951 to 31 October 2003.

2.2.1 Climatic and hydrologic data

Daily meteorological data (maximum, minimum and mean air temperature, precipitation amounts, relative humidity, sunshine and total cloud cover durations) were provided by the German Weather Service (DWD) and the Czech Hydrometeorological Institute (CHMI). In the Czech part of the basin, the climate station network is less dense. The station data were corrected for inconsistencies, data gaps and inhomogeneities (Österle et al., 2006, 2012).

The soil map was generated by merging the German soil map “BUEK 1000” provided by the Federal Institute for Geosciences and Natural Resources (BGR) and the FAO-UNESCO soil map for the Czech part. The resolution and quality of the soil maps are different, which may influence the results of the rainfall-runoff model.

The land use information is taken from the CORINE 2000 land cover data set of the European Environment Agency and is considered as static during the analysis period.

Discharge data were provided by various German water authorities and the Global Runoff Data Centre (GRDC). 114

gauging stations (Fig. 1, dots), including a large number of nested catchments, were used for flood identification. While the selected gauges are densely and approximately equally distributed in the German part, six gauges were available in the Czech part. Catchment size varies between 104 km² and 131 950 km². Half of the gauges covered at least 94 % of the analysis time period. Hydrological years with more than 60 days of missing data were excluded from the flood identification resulting in between 57 (1951) and 114 (1981) gauges in the analysis. 27 discharge gauges (Fig. 1, red dots) were used for the calibration and validation of the rainfall-runoff model.

2.2.2 Scatterometer data

The remotely sensed soil water index (SWI) (Wagner et al., 1999) was provided by the Vienna University of Technology, Institute of Photogrammetry and Remote Sensing (<http://www.ipf.tuwien.ac.at/radar/ers-scat/home.htm>). Surface soil moisture is derived from the radar backscattering coefficient of the scatterometers onboard of the satellites ERS-1 (1991–2000) and ERS-2 (1995–2011). First, the backscattering coefficients are standardized to a reference incidence angle (40°) by applying a change detection method. Afterwards, they are rescaled between their minimum and maximum value to represent the driest and wettest soil moisture conditions of the topmost soil layer. The soil moisture content in the first meter of the soil (SWI) is derived by applying an exponential two-layer water model to the surface soil moisture estimates (Wagner et al., 1999). The temporal resolution of the SWI is 10 days. As soil moisture retrieval is not possible under snow and frozen soil conditions, the SWI sample times are restricted from April/May to November. 228 sample times overlap with the hydrological simulation. The distance between the pixel centroids is approximately 12.5 km (Fig. 1, crosses).

3 Methodology

Profile (layer-depth weighted average) soil moisture content is simulated with a rainfall-runoff model. For the comparability of the profile soil moisture content of different soil types, the values are standardized by the field capacity of each soil type. In the following, the standardized profile soil moisture content is termed soil moisture index (SMI). PCA is used to map daily spatial SMI onto specific spatial patterns that express large parts of the spatial variability of the soil moisture series. Cluster analysis groups days of similar soil moisture patterns. In parallel, flood start days are derived from observed discharge time series and flood prone soil moisture patterns are identified.

3.1 Flood event identification

For the investigation of the link between large-scale soil moisture patterns and flood initiation, a flood definition that takes into account the simultaneous or time shifted flooding at several gauges is required. Several flood identification methods taking the spatio-temporal coherence of flooding into account have recently been proposed by e.g. Rodda (2005), Merz and Blöschl (2003), Keef et al. (2009), Uhlemann et al. (2010) and Ghizzoni et al. (2012). We identified large-scale flood events in the Elbe River basin using an approach proposed by Uhlemann et al. (2010) as the method is non-restrictive to a certain return period and takes the simultaneous or time shifted occurrence of peak discharges at many sites into account. A flood event is identified if at least one gauge within the catchment exceeds its 10-yr flood (POT). In a subsequent step, one searches for further significant peak discharges in a time frame three days in advance and ten days after the date of the POT. All peak dates around the POT are pooled into one flood event. Two flood events are independent from each other if the last occurrence of a significant peak of the previous flood and the first significant peak of the following flood event are separated by at least four days. Otherwise they are considered as one flood event. In this way, each flood event is characterized by a flood start date (first gauge showing significant peak around POT) and a flood end date (last gauge showing significant peak around POT). Furthermore, each flood is characterized by a measure of the overall event severity. The severity measure combines the overall flood extent and the flood magnitude (for details see Uhlemann et al., 2010).

To explore the link between flood initiation and the occurrence of soil moisture patterns, the soil moisture patterns at the start dates of the respective flood events are examined.

3.2 Hydrological modeling

3.2.1 Model description

The continuous daily eco-hydrological model SWIM (Krysanova et al., 1998) is a conceptual, semi-distributed model based on SWAT (Arnold et al., 1993) and MATSULA (Krysanova et al., 1989). The model has three levels of spatial disaggregation: the basin (entire considered river basin), subbasins (subdivision of the basin) and hydrotops. Hydrotops are units of unique land use and soil type which are assumed to have a specific hydrological reaction. Climate input and the groundwater routine act on the subbasin scale. Daily values of relative humidity, global radiation, precipitation, mean, maximum and minimum air temperature are interpolated at the subbasin centroids. The snow routine and the soil water balance are calculated at the hydrotops and river routing at the basin scale. 1945 subbasins with a median catchment size of 33 km² (minimum 2 km², maximum 1034 km²) were implemented upstream of Wittenberge (Fig. 1).

The snow module is based on the degree-day method. Snow accumulation and melt depend on threshold air temperatures and a degree-day factor. Surface runoff is calculated with a modified version of the SCS-curve number method (Arnold et al., 1993; USDA Soil Conservation Service, 1972). In the soil routine, the soil root zone is subdivided into several soil layers in accordance with the soil profile of the specific soil type. To calculate percolation, a storage routing technique is applied on water inflow divided into slugs of 4 mm. Percolation in each layer depends on the soil water content which has to exceed field capacity and on the travel time through the layer governed by the saturated hydraulic conductivity. If the subjacent layer is saturated or if the soil temperature in a layer is below 0°C no percolation occurs. Lateral subsurface flow is a function of the remaining drainable water volume and the return flow travel time which depends on the baseflow factor and on the saturated conductivity. If the considered soil layer is saturated, water is assumed to rise to the overlying layer.

Potential evapotranspiration is estimated with the Priestley–Taylor method (Priestley and Taylor, 1972). Based on potential evapotranspiration, plant transpiration and soil evaporation are calculated separately as a function of the leaf area index according to Ritchie (1972). The actual soil evaporation is estimated on the upper 0.3 m of the soil zone. As long as the soil evaporation accumulated since the last rainfall event is below 6 mm, actual soil evaporation equals potential soil evaporation. If the soil evaporation exceeds 6 mm, actual evaporation is assumed to decay exponentially. In the case of a snow cover, the soil evaporation is retained from the snow water content. For the estimation of the actual plant transpiration, the potential water use by plants based on the root development is calculated first. Secondly, potential water use is adapted to actual water use by the soil water content and field capacity.

The groundwater module consists of a shallow and a deep aquifer. The shallow aquifer is recharged by the percolation from the bottom soil layer with an exponential delay weighting function. The return flow is the groundwater contribution to the streamflow from the shallow aquifer. The amount of seepage from the shallow aquifer to the deep aquifer and the capillary rise from the shallow aquifer back to the soil profile are estimated as linear functions of recharge and actual evapotranspiration, respectively.

Routing is calculated with the Muskingum method (Maidment, 1993, chap. 10.2.3). Surface runoff and the sum of subsurface and groundwater are routed separately.

3.2.2 Model calibration and validation

The Elbe River basin is subdivided into 27 regions, assuming homogeneous parameterization within each region (Fig. 1, red dots). Nine sensitive parameters controlling snow accumulation and melt, potential evapotranspiration, saturated hydraulic conductivity, recharge as well as discharge

Table 1. Model parameters and their calibration range.

Model parameter	Description	Calibration range
tsnowfall	Snowfall temperature [°C]	-1.0 – 2.5
tmelt	Snowmelt temperature [°C]	0.0 – 3.0
snowmeltrate	Melting rate [mm d ⁻¹]	1.0 – 4.0
thc	Correction factor for potential evapotranspiration on sky emissivity [-]	0.1 – 1.2
scor	Correction factor for saturated conductivity [-]	0.1 – 20.0
abf0	α -factor for groundwater [-]	0.0 – 1.0
bff	Baseflow factor [-]	0.0 – 2.0
roc2/roc4	Routing coefficient surface/subsurface runoff [-]	0.1 – 15.0

routing (Table 1) are calibrated over the period 1981–1989 and validated over 1951–1980 as well as 1990–2003. Parameters were estimated for each region progressively from upstream to downstream by nesting the upstream regions for which parameters are already estimated in the next step. A weighted Nash–Sutcliffe efficiency coefficient (Hundecha and Bárdossy, 2004; Nash and Sutcliffe, 1970) was used as an objective function:

$$NS = 1 - \frac{\sum w(t) (q_c(t) - q_o(t))^2}{\sum w(t) (q_o(t) - \bar{q}_o)^2}, \quad (1)$$

where q_c and q_o are simulated and observed discharge, respectively. \bar{q}_o is the average observed daily discharge in the calibration/validation period. $w(t)$ gives weight to certain parts of the hydrograph. To emphasize high flows $w(t)$ equals $q_o(t)$. The first 90 days of the simulation period are used as model initialization and excluded from the efficiency calculation.

As several parameter combinations can lead to the same model performance (Beven and Binley, 1992), a Monte Carlo uncertainty analysis is carried out to identify an ensemble of parameter sets that lead to behavioral model performances. A model performance is considered as behavioral, if an a-priori set threshold of goodness of fit is exceeded.

The Monte Carlo simulation results in i data matrices **SMI** ($m \times n$), where m is the number of observations in time (i.e. 18 993), n is the number of subbasins (i.e. 1945) and i refers to the number of behavioral parameter combinations (Monte Carlo parameter sets).

$$\mathbf{SMI}_i = \begin{pmatrix} \mathbf{SMI}_{1,1} & \cdots & \mathbf{SMI}_{1,n} \\ \vdots & \ddots & \vdots \\ \mathbf{SMI}_{m,1} & \cdots & \mathbf{SMI}_{m,n} \end{pmatrix} \quad (2)$$

In a next step, the simulated SMI is validated. The verification with soil moisture point measurements (e.g. gravimetric) is not feasible as these are highly variable over short distances, whereas satellite-based and hydrological-simulated soil moisture estimates are spatially integrated values (Parajka et al., 2006). For this reason, the temporal SMI progression is validated against the remotely sensed soil water index

(SWI) by calculating the Pearson correlation coefficient between SMI and SWI. In accordance with the area share of overlaying SWI grid points, an area-weighted average SWI is assigned to each subbasin. For each subbasin between 122 and 197 SWI estimates are available for model validation.

3.3 Principal component analysis

The standardized soil moisture simulations of the behavioral Monte Carlo runs are arranged consecutively in time reshaping **SMI** into a matrix **SMI*** of size $((i \times m) \times n)$. PCA is applied to reduce the data dimensionality of **SMI***. First, the spatial linear Pearson correlation matrix **R** ($n \times n$) of **SMI*** is calculated giving equal weight to all subbasins and Monte Carlo parameter sets. As **R** is square and symmetric and therefore diagonalizable, one can identify the eigenvectors **u** (specific spatial patterns expressing large parts of the spatial SMI variability) and the eigenvalues $\text{diag}(\lambda)$ of the matrix **R**:

$$\mathbf{R}\mathbf{u} = \lambda\mathbf{u}, \quad (3)$$

by solving

$$(\mathbf{R} - \lambda\mathbf{I})\mathbf{u} = 0, \quad (4)$$

where **I** is the $(n \times n)$ identity matrix. The eigenvectors **u** are sorted in decreasing order of their corresponding eigenvalues λ as the eigenvalue λ_k is a measure of the explained variance ev_k of the corresponding principal component PC_k :

$$ev_k = \frac{\lambda_k}{\sum_{k=1}^n \lambda_k}. \quad (5)$$

Hence, the leading first eigenvector points in the direction of the highest variance of **SMI*** and the next eigenvector explains the subsequent highest variance with the condition being orthogonal to the already identified eigenvector.

The PCs are obtained by projecting the standardized (zero mean, unit variance) data matrix of **SMI*** onto the eigenvector \mathbf{u}_k .

$$PC_k = \mathbf{SMI}_{std}^* \mathbf{u}_k \quad (6)$$

As **SMI*** compromises all behavioral Monte Carlo runs, uncertainty bounds of each PC_k are obtained by decomposing PC_k of size $((i \times m) \times 1)$ into i time series of length m .

The PCs are tested for significance using the rule-N approach (Overland and Preisendorfer, 1982). The calculated normalized eigenvalues are compared against the normalized eigenvalues of a random Gaussian matrix. Those leading normalized eigenvalues that are higher than the 95th percentile of the simulated random eigenvalues (1000 Monte Carlo runs) are treated as significantly different from a random field. For further details on PCA see e.g. Hannachi et al. (2007), Jolliffe (1986) or Preisendorfer (1988).

3.4 Cluster analysis

The hierarchical Ward cluster algorithm (Ward Jr., 1963) is implemented on the leading PCs to identify days of similar soil moisture patterns. At the beginning, every day D represents a single cluster. At each analysis step, the union of all possible cluster pairs is considered and the cluster pair t that offers the smallest increase in variance V is merged.

$$V_t = \sum_D \sum_k |PC_{tDk} - \bar{PC}_{t,k}|^2, \quad (7)$$

where PC_{tDk} denotes the value of the k -th PC at day D belonging to cluster t . The algorithm merges the days consecutively until all days are united in a single cluster.

Including all behavioral Monte Carlo runs, the extent of a single PC is $((i \times m) \times 1)$ i.e. $((38 \times 18\,993) \times 1)$. Due to limits in the computer capacity, it is not feasible to accomplish a cluster analysis directly on the leading PCs. Thus, we execute the cluster analysis in two steps. First, cluster analysis is carried out on the leading PCs of each behavioral Monte Carlo run separately. To merge the cluster results of the behavioral Monte Carlo runs, a second cluster analysis is applied on the cluster centroids (median of respective cluster members) of the behavioral Monte Carlo runs. It is assumed that the number of clusters in the first and second cluster step is the same.

Depending on the parameterization of a particular Monte Carlo run, a single day D may have different pattern characteristics and is thus assigned to different clusters t . Subsequently, each day D is assigned to the cluster with the highest probability of occurrence expressed as p_D :

$$p_D = \max \left(\frac{|D \in t|}{i} \right). \quad (8)$$

In order to estimate the influence of model parameterization on the cluster t , the median p_D of all days belonging to a specific cluster t is calculated, which defines the probability of cluster membership p_t :

$$p_t = \text{median}(p_D \in t). \quad (9)$$

Thus, a small probability of cluster membership p_t indicates a strong influence of the model parameterization on the cluster assignment, while a large probability of cluster membership p_t indicates a weak influence.

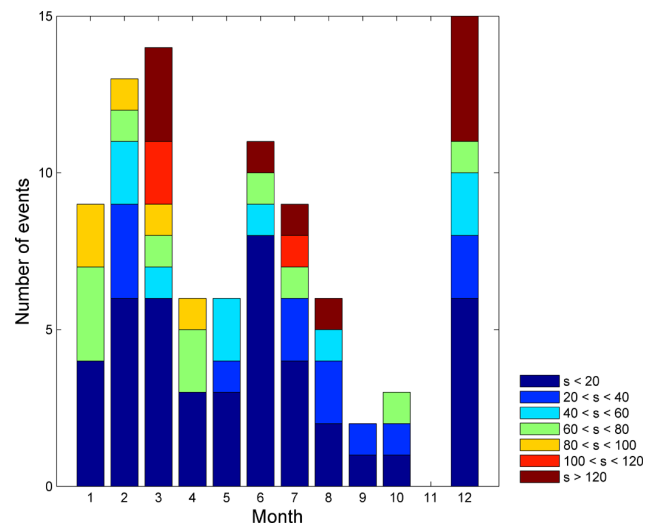


Fig. 2. Number of flood start dates per month and severity class s .

To determine the number of leading PCs in the clustering as well as the number of clusters (Eq. 7), a sensitivity analysis is conducted varying the number of PCs and the number of clusters. Based on p_t (Eq. 9) the optimum PC-cluster combination is selected.

4 Results

4.1 Identified flood events

From the observed discharge time series, 94 flood events are identified, out of which 60% are winter (November–April) events. Figure 2 displays the flood events separated by the month of the flood start date and the severity class s . In February, March, June and December more than ten flood events are initiated. September and October have by far the lowest number of flood initiation. In November, no flood events are initiated. High severities ($s > 100$) are found in the winter months December as well as March and in the summer months June to August. The severe events are not restricted to the months with the highest number of flood initiation. As the severity is a combined measure of flood magnitude and extent (affected river network), one has to take their respective influence into account. Winter events are characterized by a large spatial extent of minor magnitudes. In contrast, summer events are either characterized by a small spatial extent of very few extreme magnitudes or by a large spatial extent of miscellaneous magnitudes.

4.2 Hydrological modeling

The application of the Monte Carlo approach resulted in 38 behavioral parameter sets. In the calibration period, all gauges have a median weighted Nash–Sutcliffe efficiency between 0.55 and 0.8. The performance difference between the

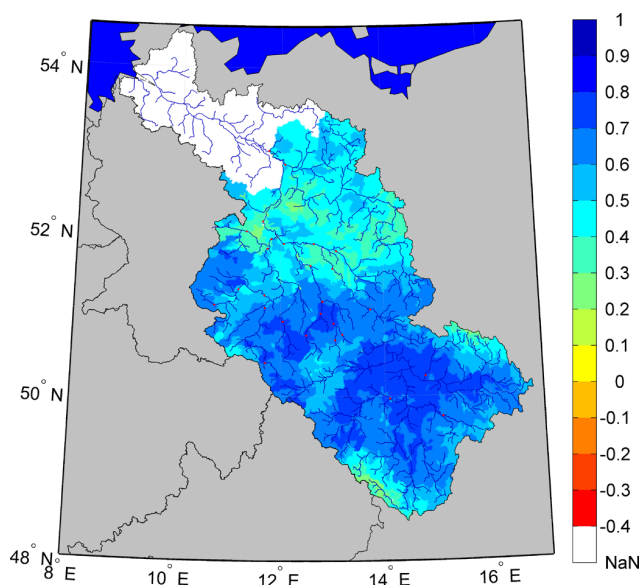


Fig. 3. Pearson correlation coefficient (significant level 5 %) between the median simulated SMI and the remotely sensed SWI.

behavioral Monte Carlo parameter sets is negligible. Their median weighted Nash–Sutcliffe efficiency ranges between 0.71 and 0.74. In the validation period, the gauges median weighted Nash–Sutcliffe efficiency ranges between 0.53 and 0.81 (1951–1980) as well as 0.26 and 0.87 (1990–2003). Gauge Havelberg has by far the lowest efficiency which can be attributed to various lakes and strong anthropogenic modifications (mining) not represented in the model. Due to the chosen model performance measure, the calibration puts more emphasis on high flows compared to low flows. As a consequence, runoff volume is in the median overestimated by 33 % in the calibration period and by 28 % in the validation period 1951–1980 (40 % in 1990–2003).

For the validation of the SMI, the Pearson correlation coefficient (significant level 5 %) between the median simulated SMI and the remotely sensed SWI was calculated (Fig. 3). Except for a few local spots, basin-wide high correlations are observed. The median correlation is 0.57, the difference between the 25th and 75th quantile is 0.2. The individual examination of the Monte Carlo parameterizations leads to the same findings. Hence, the simulated temporal progression of the SMI is well represented in the hydrological model.

4.3 Soil moisture pattern classification

4.3.1 Principal component analysis

The leading 43 out of 1945 PCs are significantly different from a random field explaining 97.66 % of the total variance. The leading four PCs and their corresponding eigenvectors are displayed in Fig. 4. The eigenvectors indicate geographic regions of simultaneous anomalies in the SMI. The minimum

and maximum values of the PCs correspond to the parameter uncertainty introduced by the rainfall-runoff model.

The first PC explains 75.71 % of the total variance and has a seasonal behavior. The influence of the parameter uncertainty on the PC is small. The corresponding loadings show a low spatial variability across the catchment (Fig. 4, top). The subsequent PCs have a damped and lagged seasonal behavior and their loadings are spatially heterogeneous. The second PC explains 8.60 % of the total variance, seasonality is still visible. Compared to the first PC, the influence of the parameter uncertainty on the PC increased, although the general behavior is the same. Its corresponding loadings show a north-south partition. The German part of the river basin has positive loadings excluding the upstream areas of the Saale and Mulde, whereas the Czech part of the river basin has negative loadings. The third PC, explaining 1.87 % of the total variability, has no apparent periodic behavior. High positive loadings are found in the Saale region and the mountainous area of the catchment. The fourth PC shows the highest influence of parameter uncertainty of the presented PCs. The explained variance is 1.49 %. The loadings are positive in large parts of Saale region, slightly positive in the central Czech part of the catchment and highly negative in the downstream Havel region.

4.3.2 Cluster analysis

Up to 15 of the leading PCs were chosen as variables in the cluster analysis to identify days of similar soil moisture patterns. This implies that between 75.71 % (first PC) and 94.10 % (all leading 15 PCs) of explained variance are involved in the clustering. Applying the variance criteria of Ward (Eq. 7), the weight assigned to a particular PC depends on its explained variance, as the individual PCs are not standardized, i.e. the first PC has the highest weight.

The probability of cluster membership p_t (Eq. 9), which evaluates the influence of model parameterization, was used to select an optimum PC-cluster combination. Figure 5 (left) displays the median p_t of different PC-cluster combinations. A division in two clusters leads in the median to well distinguishable clusters independent of model parameterization (median $p_t \sim 1$). The median p_t decreases up to a division in approximately 12 clusters and remains stable for a higher number of clusters. In general, the median p_t is independent of the number of PCs in the cluster analysis. An exception is the clustering of three or less PCs. In this case, the included variability is insufficient to derive a high number of distinguishable clusters as indicated by a low median p_t . In the following, the number of PCs in the cluster analysis is fixed to four since the variability (86.67 % of explained variance) is sufficient to derive a high number of distinguishable clusters.

Henceforward, clustering the leading four PCs, Fig. 5 (right) displays the distribution of p_t for different cluster numbers. In almost all cases, clusters with a very high p_t

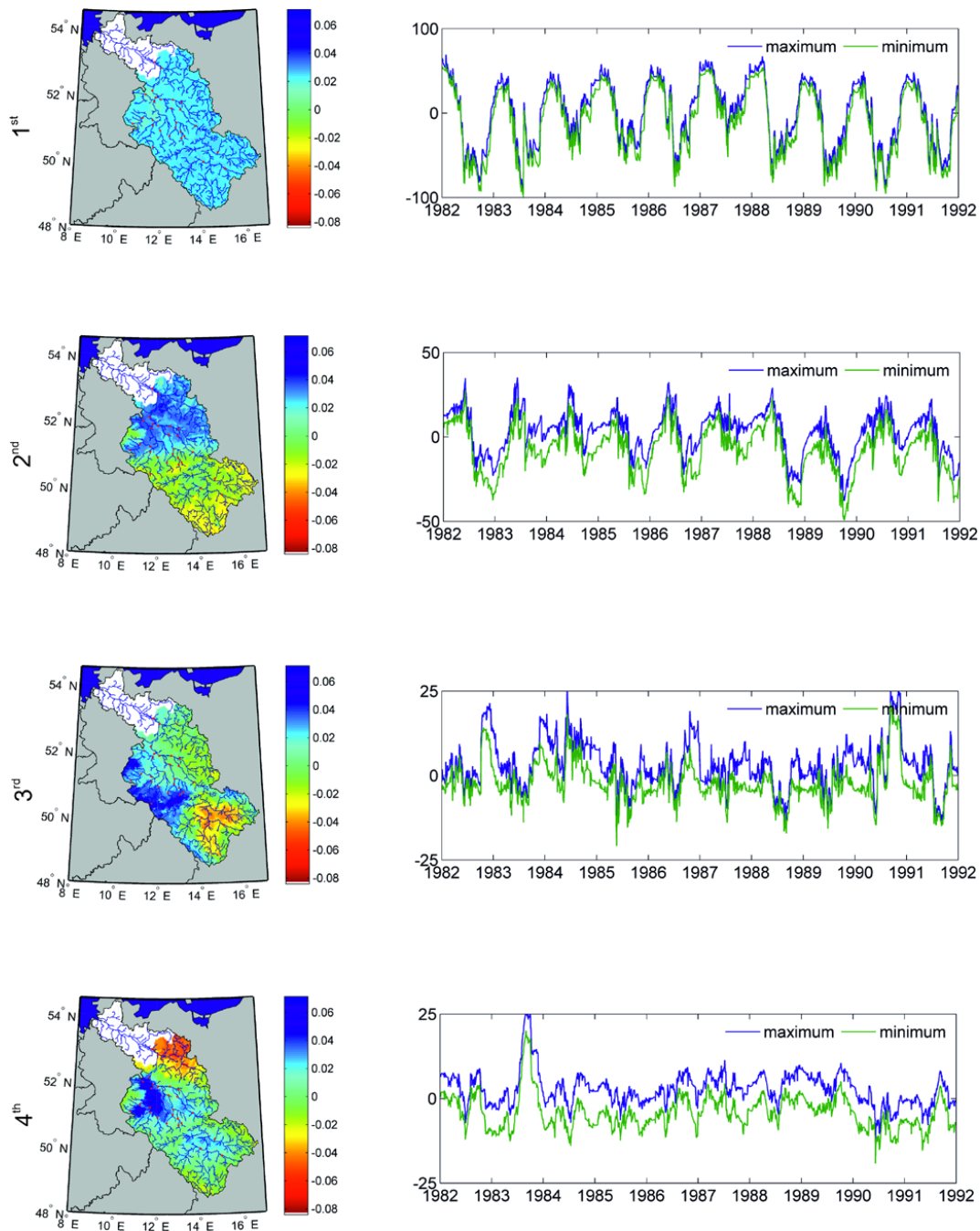


Fig. 4. Eigenvectors (left) and their corresponding PCs (right) of the leading four PCs. PCs are displayed for the sub-period 1982–1991. Minimum and maximum values correspond to the parameter uncertainty introduced by the rainfall-runoff model.

and thus independent of model parameterization and clusters with low p_t (< 0.5) and thus dependent on model parameterization can be found. In a subsequent step, the analysis is restricted to ten clusters, since the median p_t peaks for ten clusters (0.74) and decays for more clusters.

The SMI patterns of the cluster centroids, when separating the leading four PCs into ten clusters, are presented in Fig. 6. Cluster one, seven and eight show the driest soil moisture

patterns. Between 87.8 % and 97.5 % of the catchment area has an SMI below 0.6. Clusters two and ten have intermediate SMI. Around 72 % of the catchment area has an SMI below 0.6. High SMI values are mainly restricted to the mountainous areas. Around 4 % of the catchment area has an SMI above 0.8. The remaining clusters change consecutively into highly saturated soil moisture patterns. Initially, a high SMI is restricted to the mountainous areas and intermediate values

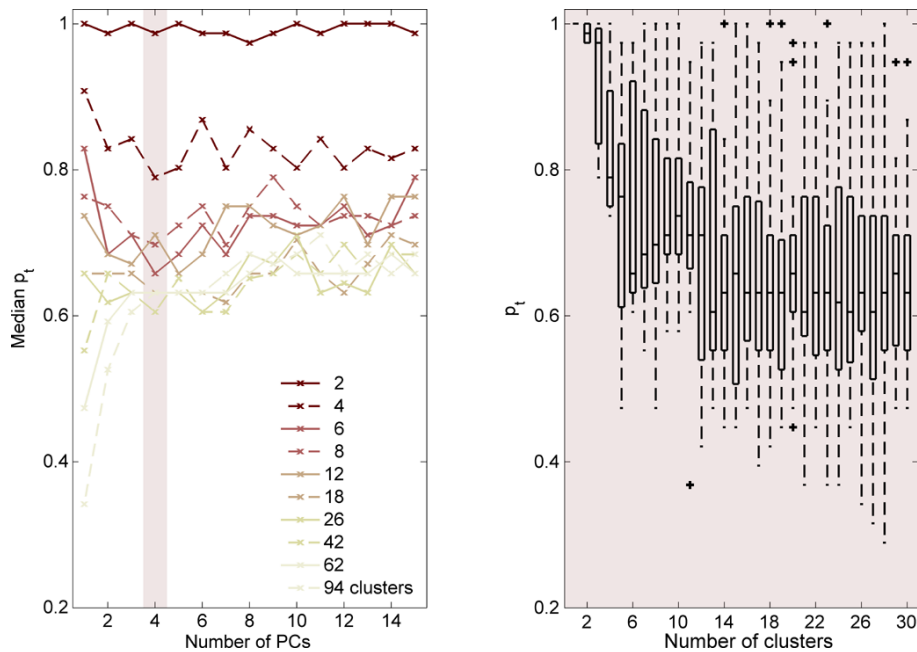


Fig. 5. Median probability of cluster membership p_t of different PC-cluster combinations (left). PC-cluster combinations with a small median p_t are strongly influenced by model parameterization. Distribution of p_t for different numbers of clusters when clustering the leading four PCs (right).

dominate in the mid- and lowlands (pattern six). The SMI rises in the upstream part of the basin (pattern five). The northern part of the basin gets wetter (pattern four) and the SMI in the upstream part of the basin increases additionally (pattern three). Here, 28.4 % of the catchment area has an SMI above 0.8. Finally, cluster nine shows high SMI over the entire basin. 43.8 % of the catchment area has an SMI above 0.8 and 90.1 % of the catchment area has an SMI above 0.6.

Table 2 displays the statistics of the ten clusters. Each day is represented 38 times, corresponding to the number of behavioral parameter sets. Cluster nine is the biggest cluster, containing 24.1 % of total days. Clusters two and seven are the smallest clusters, containing around 5 % of total days. The probability of cluster membership p_t ranges between 1.0 and 0.58. Cluster nine has the highest p_t . Clusters two and six have the lowest p_t . Here, in the median 22 out of 38 parameter sets assign a respective day to the clusters. In order to indicate the day-to-day variability of the classification, the median persistence of each pattern type/cluster is calculated. Cluster nine has the highest persistence. In the median, the pattern persists for eleven days (average of 43 days). With a median duration of seven days, clusters three and seven have the second highest persistence. The other pattern types have a median duration of either four or five days. The monthly frequencies of the different pattern types are presented in Fig. 7 (top). The frequencies express the relative occurrence of a particular pattern type in each month. Seasonal differences between the pattern types are visible. For instance, cluster three and nine are winter patterns. Cluster seven occurs in

summer. Cluster four predominates in April, whereas cluster one predominates in June. In winter and in particular at the beginning of the year, clusters three and nine are the dominant patterns, whereas in summer/autumn the occurring clusters are more various.

4.4 Soil moisture patterns and their relationship to flood initiation

In the following, the soil moisture pattern types are characterized according to their relationship with flood initiation. Since the flood start days are considered for all parameter sets separately, 38 times 94 flood start days are included in the analysis. Cluster nine comprises the highest percentage (52.9 %) of flood start days (Table 2). Clusters eight and ten contain less than 2 % of flood start days each. Cluster seven contains nearly no flood start days. The frequency of the flood start days within each cluster expresses how often the respective pattern type can be related to flood initiation. Cluster nine has the highest frequency of flood start days. The frequency of flood initiation is 1.08 %. The frequency of flood start days in cluster eight is one order of magnitude lower. In comparison with no clustering, flood start days are accumulated in cluster nine and the remaining clusters account for relatively few flood start days.

The relative frequency of flood start days within a respective month and pattern type are presented in Fig. 7 (bottom). In summer, the highest relative frequencies of flood start days can be found and a multiplicity of pattern types are related to

Table 2. Cluster statistics.

Cluster #	1	2	3	4	5	6	7	8	9	10	no clustering
Days [%]	7.3	5.3	13.6	9.3	7.9	10.4	5.2	8.0	24.1	8.9	100
p_t [-]	0.82	0.58	0.68	0.71	0.68	0.58	0.79	0.82	1.0	0.76	1.0
Median persistence [days]	5	4	7	5	5	4	7	5	11	5	1
Flood start days [%]	4.8	4.8	11.0	8.2	6.7	8.2	0.2	1.9	52.9	1.3	100
	(172)	(172)	(393)	(292)	(241)	(293)	(7)	(69)	(1888)	(45)	(38 × 94)
Frequency of flood start days [%]	0.33	0.45	0.40	0.44	0.43	0.39	0.02	0.12	1.08	0.07	0.49
Median flood severity [-]	1.9	4.6	37.1	27.5	20.8	7.4	1.0	21.7	46.1	4.6	25.68

flood initiation. In winter, cluster three and in particular cluster nine, both characterized by high soil moisture across the entire basin, are related to flooding. Here, the relative frequency of flood start days is approximately constant from January until April/May. Although the two clusters have approximately the same seasonal distribution (Fig. 7, top), cluster nine is primarily relevant for flood initiation in winter, whereas cluster three is of primary importance in summertime flood initiation. In June, cluster three has the highest relative frequency of flood start days of all clusters. This seasonality shift between pattern frequency in general (Fig. 7, top) and pattern flood initiation (Fig. 7, bottom) also occurs in the case of clusters four and two.

Beside the pattern type frequency and relationship to flood initiation, the pattern type persistence three days in advance and after the flood start date was investigated (not shown). On the one hand, the patterns after the flood start date are more persistent than in the days ahead of the flood start date. On the other hand, there is a clear difference in the pattern persistence between summer and winter events. In wintertime, the soil moisture pattern types are persistent both in advance and after the flood start date. Occasionally, the patterns shift from cluster three into cluster nine and vice versa in the days ahead of the flood start date. In summer, either pattern type three is persistent in advance of the flood start date or a continuous wetting (transformation from pattern six, five or four) occurred. The summer patterns of low SMI are either persistent or transform into wetter patterns after the flood start date.

Finally, the relationship between the soil moisture patterns types and flood severity is analyzed (Table 2). For high SMI (patterns three, four and nine) the median flood severity increases with the degree of saturation. For dry soil moisture conditions this is not the case. Cluster eight has a median flood severity of 21.7, whereas the wetter pattern six has a median flood severity of 7.4.

5 Discussion

5.1 Soil moisture pattern classification

Large-scale soil moisture dynamics were simulated with a hydrologic model. As the model was calibrated on discharge, an integrated measure of catchment processes, the spatially distributed processes within the gauged catchment may not be well represented and small-scale patterns may be a calibration artifact. In the present study, especially the temporal SMI progression was of relevance since the PCA was based on the spatial correlation matrix (Eq. 3). The validation of the simulated SMI with the remotely sensed SWI proved the basin-wide reliability of the temporal SMI progression. As there is no obvious difference in the correlations between the Czech and the German part of the basin, an impact of the different soil maps and climate station densities on the temporal progression of simulated soil moisture can be excluded. Applying a different hydrologic model in Austria, Parajka et al. (2006) obtained considerably lower correlation coefficients of root zone soil moisture (median of 0.07 (1991–1995) and 0.12 (1996–2000)). As soil and land use properties are included in the model, there is a physical meaning behind the soil moisture patterns. For this reason, we did not interpolate the SMI on a regular grid as done by other studies (e.g. Perry and Niemann, 2007; Kim and Barros, 2002). Rather, we conducted the PCA at the subbasin scale with the drainage divides as natural boundaries.

The large amount of total variability explained by the first PC indicates that there is one dominant mode controlling the temporal soil moisture variability. This temporal mode has a seasonal behavior with maximum values in winter and spring and minimum values in summer (Fig. 4, top right). These seasonal soil moisture changes are mainly attributable to seasonal changes in evapotranspiration, leading to soil moisture depletion in summer and rise in winter and spring (Parajka et al., 2010). Additionally, snowmelt may have an impact. All subbasins respond in the same direction in terms of seasonality, indicating considerable similarity in the processes controlling soil moisture variability (Fig. 4, top left). Heterogeneous hydrological processes and catchment properties are overruled by seasonality. Perry and Niemann (2007) reported

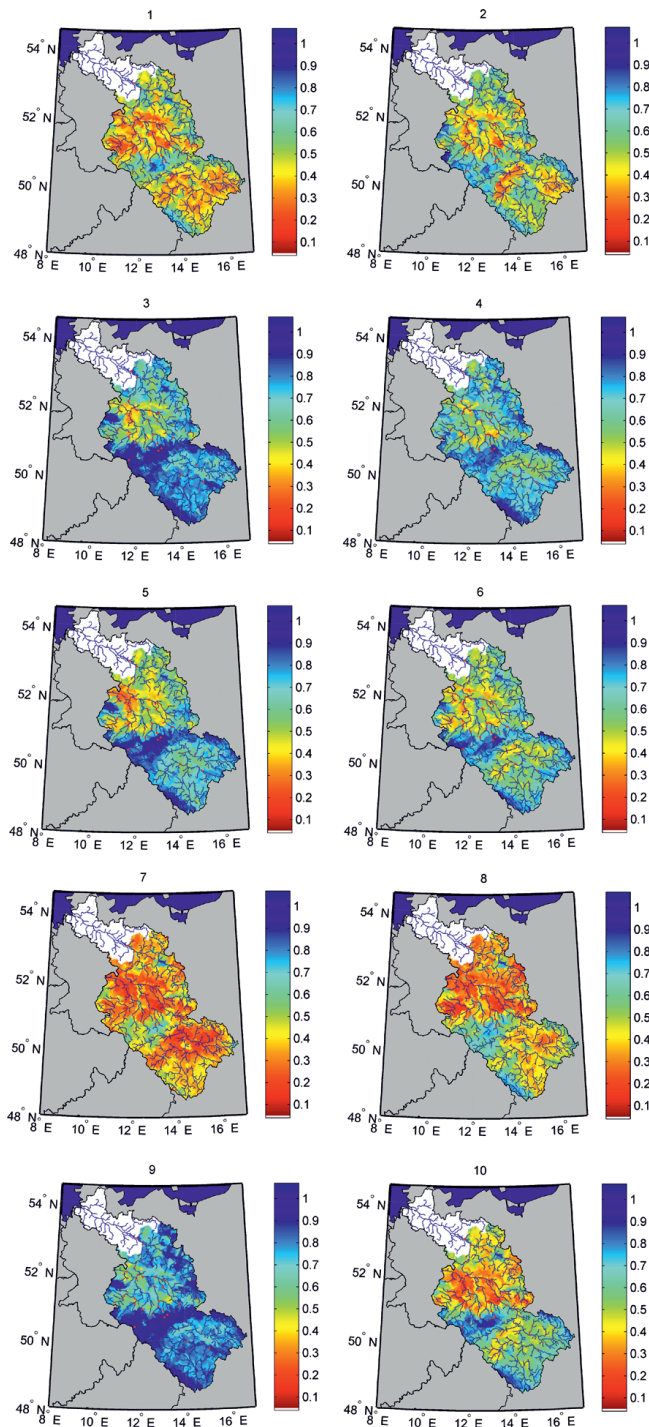


Fig. 6. Soil moisture index (SMI) patterns of cluster centroids.

similar findings for the leading PC. Limiting the analysis to a smaller area receiving evenly distributed precipitation would possibly identify PCs attributable to single precipitation events. As the loadings are not only positive across the catchment but also approximately of the same magnitude, the first PC is a measure of the catchment average SMI. In

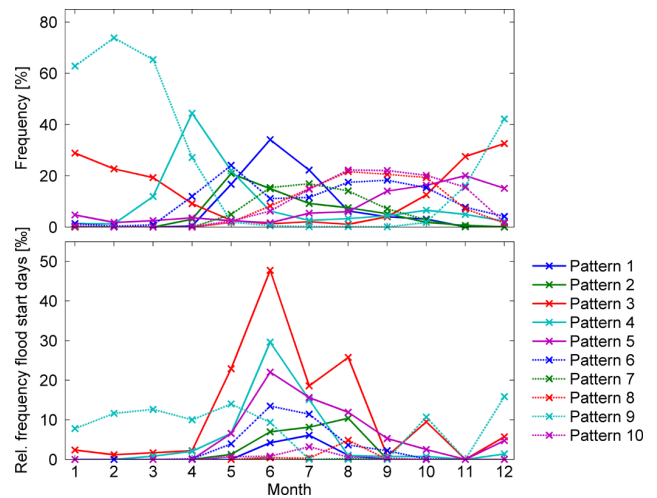


Fig. 7. Frequency of occurrence per month of different soil moisture pattern types [%] (top). Relative frequency of flood start days per month and respective soil moisture pattern type [%] (bottom).

contrast, the subsequent PCs are a measure of the disparity, describing local variations departing from the regional value. Their damped and lagged temporal progression might be attributable to the antecedent soil moisture conditions, i.e. the previous spatial distribution of precipitation.

The PCA showed that the retrieved soil moisture patterns have a stronger signal than the parameter uncertainty introduced by the rainfall-runoff model (Fig. 4, right). The first PC has the smallest relative difference between the members of the parameter ensemble and is least influenced by model parameterization. The relative differences between the ensemble members are higher for the subsequent PCs revealing a greater impact of model parameterization on spatially heterogeneous hydrological processes, as e.g. infiltration and soil storage, than on the seasonal progression of soil moisture.

The cluster analysis identified days of similar soil moisture patterns. Each soil moisture pattern type can be attributed to a pattern frequency, a pattern persistence and seasonal characteristics (Table 2, Fig. 7) in the same manner as in weather type classifications (e.g. Philipp et al., 2010). The probability of cluster membership p_t yielded an objective choice of the number of PCs as well as clusters with respect to the parameter uncertainty of the rainfall-runoff model. Pattern nine characterized by high soil moisture content over the entire catchment is least dependent on model parameterization (highest p_t) and most persistent. Independent of model parameterization, any additional rainfall will be transformed into runoff without leading to alteration of the pattern type. Dry SMI patterns have a high p_t too. Likewise, the absence of rainfall does not lead to an alteration of the pattern type. Intermediate SMI patterns show the lowest p_t and are comparatively less persistent, as these patterns are a transitional stage to either wetter or dryer soil moisture conditions.

5.2 Soil moisture patterns and their relationship to flood initiation

In flood frequency analysis and design value estimation, flood events are defined for one particular gauge. An extreme value distribution is fitted to annual maxima series or a peak-over-threshold series of observed discharge. In the present study, the flood identification differed from this commonly applied approach. Instead, a flood identification method that takes into account the large-scale response has been adopted (Uhlemann et al., 2010). This is a more suitable approach in terms of the linkage between floods and large-scale soil moisture pattern types. Among the 94 detected flood events, the most severe events are well documented in terms of hydro-meteorological conditions, e.g. the flood events of August 2002 (Engel, 2004; Ulbrich et al., 2003a, b), July 1954 (Hauptamt für Hydrologie, 1954; Boer et al., 1959), or December 1974/January 1975 (Schirpke et al., 1978). Nevertheless, the flood event set may be incomplete due to ungauged small catchments receiving convective rainfall, as well as the requirement of an observed 10-yr flood. Furthermore, the flood event set may be biased towards flooding in the German part of the Elbe catchment, as only six gauges were available in the Czech Republic.

Previously, Uhlemann et al. (2010) highlighted the occurrence of summer flooding in the Elbe Basin. However, when analyzing trans-basin German floods, they found that the severe events were limited to the winter period. Our study, which is limited to flood identification in the Elbe, results in events of high severity ($s > 100$) both in winter and summer. Analyzing annual maximum flood series in Austria, Merz and Blöschl (2009) did not detect a dependence between the flood moments and seasonality. Instead, mean annual flood flows were positively correlated with average antecedent rainfall as a surrogate for the catchment soil moisture state (Merz and Blöschl, 2009). Comparing the soil moisture pattern types and their respective median flood severity (Table 2), the results are heterogeneous and reflect different flood-generating processes. On the one hand, the median flood severity increased for patterns of high SMI. In this case, the catchment storage capacity is exceeded, e.g. due to long-term low-intensity rainfall, and any additional rainfall results in a runoff increase. In addition, if most parts of the catchment are saturated, a larger number of gauges may be flood affected in case of synoptic rainfall, as shown on the pan-European scale by Prudhomme and Geneviev (2011). In their regional flood typology, Merz and Blöschl (2003) termed this type long-rain flood. On the other hand, the relatively dry soil moisture pattern eight shows high median severity too. This is attributable to high-intensity rainfall on relatively dry catchment conditions, which may lead to flash floods (Merz and Blöschl, 2003). The flood-generating processes are revealed in the seasonality of the soil moisture pattern types. In winter, flooding is related to high soil moisture content in the entire basin (cluster nine). The respective

soil moisture pattern is persistent before and after the flood start date, indicating flooding either due to long-term low-intensity rainfall (long-rain flood) or saturated soils due to snowmelt (snowmelt flood). In summer, the flood-initiating soil moisture patterns are more variable and less persistent. On the one hand, large-scale flooding is observed. Long lasting rainfall leads to high soil moisture content over large parts of the basin (cluster three). On the other hand, convective events (flash floods) characterized by relatively dry catchment conditions (cluster eight) occur (IKSE, 2005).

In addition to soil moisture, other patterns, in particular precipitation, are relevant for flood initiation (Merz and Blöschl, 2008, 2009; Brocca et al., 2008; Parajka et al., 2010; Marchi et al., 2010). This is indicated by the seasonal cluster distribution and the deviating seasonality of the flood start days inside the cluster (Fig. 7, e.g. cluster three). Furthermore, extending the pattern classification approach to e.g. precipitation patterns, the role of saturated soils limited to parts of the catchment when receiving rainfall may be pronounced.

6 Conclusions

Flood generation and magnitude are the result of a complex interaction between the meteorological situation and pre-event hydrological catchment conditions. The impact of catchment conditions on floods and flood severity is expected to depend on various factors, such as season or flood type (e.g. snowmelt flood, long-rain flood, flash flood) and is especially difficult to decipher in large river catchments with catchment-internal variation in flood generation. To date, no studies are available that attempt to understand the space-time behavior of large-scale hydrological catchment conditions and link them to flood initiation. As a step in this direction, we propose classifying the hydrological catchment conditions and link flood occurrence to large-scale catchment state patterns. This approach is complementary to the widespread classification of circulation patterns in meteorology. As soil moisture is a key variable of hydrological catchment conditions, model-simulated soil moisture was used to answer two questions: what is the dominant space-time behavior of soil moisture at the regional catchment scale? Are there soil moisture patterns that are related to large-scale flood initiation? By applying PCA, the dimensionality of the soil moisture data was reduced to four PCs representing 86.67 % of the total soil moisture variability. The seasonally wetting and drying of the catchment represented by the first PC is the dominant mode, whereas the successive PCs describe spatially heterogeneous catchment processes. Cluster analysis assessed the similarity in daily soil moisture distribution within the catchment, and assigned each day of the investigation period to one of ten soil moisture pattern types. In parallel, 94 start days of large-scale flood events were identified and enabled a probabilistic linkage between

flood occurrence and large-scale soil moisture patterns. The following conclusions can be drawn:

1. Different soil moisture patterns are not equally associated with flood occurrence.
2. Patterns with catchment-wide high soil moisture accumulate the majority of flood start days.
3. Flood-initiating soil moisture patterns vary seasonally.
4. Seasonality of soil moisture pattern frequency and seasonality of soil moisture pattern flood initiation are not identical.
5. Occurrence of a certain soil moisture pattern does not necessarily lead to flood initiation, but the probability of occurrence of a large-scale flood may be increased.

While these results underline the importance of catchment state for flood initiation and severity, (4) and (5) indicate that, beside soil moisture, other patterns are relevant for flood initiation. Therefore, future work will extend the pattern classification approach not only to circulation patterns but also to snow. A combination of hydro-meteorological pattern types would enable to quantify the interaction of patterns of hydrological catchment conditions and meteorological conditions on flood initiation and magnitude.

Acknowledgements. The first author acknowledges financial support by the AXA Research Fund project “The AXA project on Large-Scale European Flooding under climate change”. We would like to thank Steffi Uhlemann for her assistance in the flood identification as well as the Potsdam Institute for Climate Impact Research (PIK) for making available the climate input data, the source code of the SWIM model and helping in setting up the model.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: F. Pappenberger

References

- Alila, Y. and Mtiraoui, A.: Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques, *Hydrol. Process.*, 16, 1065–1084, 2002.
- Apipattanavis, S., Rajagopalan, B., and Lall, U.: Local polynomial-based flood frequency estimator for mixed population, *J. Hydrol. Eng.*, 15, 680–691, 2010.
- Arnold, J. G., Allen, P. M., and Bernhardt, G.: A comprehensive surface-groundwater flow model, *J. Hydrol.*, 142, 47–69, 1993.
- Bárdossy, A. and Filiz, F.: Identification of flood producing atmospheric circulation patterns, *J. Hydrol.*, 313, 48–57, 2005.
- Beven, K. J. and Binley, A. M.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, 1992.
- Boer, W., Schubert, H., and Wilser, O.: Das Sommerhochwasser der Elbe im Juli 1954, Akademie-Verl. Berlin, 1959.
- Brocca, L., Melone, F., and Moramarco, T.: On the estimation of antecedent wetness conditions in rainfall-runoff modelling, *Hydrol. Process.*, 22, 629–642, 2008.
- Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R.: Antecedent wetness conditions based on ERS scatterometer data, *J. Hydrol.*, 364, 73–87, 2009.
- CORINE European Environment Agency: Land Cover Data Set, 2000.
- Engel, H.: The flood event 2002 in the Elbe river basin causes of the flood, its course, statistical assessment and flood damages, *Houille Blanche*, 6, 33–36, 2004.
- Ghizzoni, T., Roth, G., and Rudari, R.: Multisite flooding hazard assessment in the Upper Mississippi River, *J. Hydrol.*, 412–413, 101–113, 2012.
- Hannachi, A., Jolliffe, I. T., and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, *Int. J. Climatol.*, 27, 1119–1152, 2007.
- Hattermann, F. F., Wattenbach, M., Krysanova, V., and Wechsung, W.: Runoff simulations on the macroscale with the ecohydrological model SWIM in the Elbe catchment-validation and uncertainty analysis, *Hydrol. Process.*, 19, 693–714, 2005.
- Hauptamt für Hydrologie: Witterung und Wasserstände im Juli 1954, *Wasserwirtschaft - Wassertechnik*, Jahrgang 4, 351–652, 1954.
- Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J. Hydrol.*, 292, 281–295, 2004.
- Ibrahim, H. M., and Huggins, D. R.: Spatio-temporal patterns of soil water storage under dryland agriculture at the watershed scale, *J. Hydrol.*, 404, 186–197, 2011.
- IKSE: Die Elbe und ihr Einzugsgebiet, Ein geographisch-hydrologischer und wasserwirtschaftlicher Überblick, 262, 2005.
- Jacobbeit, J., Philipp, A., and Nonnenmacher, M.: Atmospheric circulation dynamics linked with prominent discharge events in Central Europe, *Hydrolog. Sci. J.*, 51, 946–965, 2006.
- Jawson, S. D. and Niemann, J. D.: Spatial patterns from EOF analysis of soil moisture at a large scale and their dependence on soil, land-use, and topographic properties, *Adv. Water Resour.*, 30, 366–381, 2007.
- Jolliffe, I. T.: Principal component analysis, Springer Series in Statistics, Springer, New York, 1986.
- Keef, C., Svensson, C., and Tawn, J. A.: Spatial dependence in extreme river flows and precipitation for Great Britain, *J. Hydrol.*, 378, 240–252, 2009.
- Kim, G. and Barros, A. P.: Space-time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data, *Remote Sens. Environ.*, 81, 393–403, 2002.
- Korres, W., Koyama, C. N., Fiener, P., and Schneider, K.: Analysis of surface soil moisture patterns in agricultural landscapes using Empirical Orthogonal Functions, *Hydrol. Earth Syst. Sci.*, 14, 751–764, doi:10.5194/hess-14-751-2010, 2010.
- Krysanova, V., Meiner, A., Roosaare, J., and Vasilyev, A.: Simulation modelling of the coastal waters pollution from agricultural watershed, *Ecol. Model.*, 49, 7–29, 1989.
- Krysanova, V., Müller-Wohlfeil, D.-I., and Becker, A.: Development and test of a spatially distributed hydrological/water quality

- model for mesoscale watersheds, *Ecol. Model.*, 106, 261–289, 1998.
- Maidment, D. R.: *Handbook of Hydrology*, McGraw-Hill, Inc. New York, 1993.
- Marchi, L., Borga, M., Preciso, E., and Gaume, E.: Characterisation of selected extreme flash floods in Europe and implications for flood risk management, *J. Hydrol.*, 394, 118–133, 2010.
- Merz, R. and Blöschl, G.: A process typology of regional floods, *Water Resour. Res.*, 39, 1340, doi:10.1029/2002wr001952, 2003.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007wr006744, 2008.
- Merz, R. and Blöschl, G.: Process controls on the statistical flood moments – a data based analysis, *Hydrol. Process.*, 23, 675–696, 2009.
- Merz, R., Blöschl, G., and Parajka, J.: Spatio-temporal variability of event runoff coefficients, *J. Hydrol.*, 331, 591–604, 2006.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models. Part I. A discussion of principles, *J. Hydrol.*, 10, 280–290, 1970.
- Norbiato, D., Borga, M., Merz, R., Blöschl, G., and Carton, A.: Controls on event runoff coefficients in the eastern Italian Alps, *J. Hydrol.*, 375, 312–325, 2009.
- Österle, H., Gerstengarbe, F. W., and Werner, P. C.: Ein neuer meteorologischer Datensatz für Deutschland, 1951–2003, in: *Proceedings der 7. Deutschen Klimatagung, Klimatrends: Vergangenheit und Zukunft*, Meteorologisches Institut der Ludwig-Maximilians-Universität, München, 3 pp., 2006.
- Österle, H., Schmidt, S., Hauf, Y., and Wechsung, F.: Erstellung und Prüfung eines synoptischen meteorologischen Tagesdatensatzes von 1951 bis 2003 für den tschechischen Teil des Elbe-Einzugsgebietes, in: *Die Elbe und ihr Einzugsgebiet im globalen Wandel*, edited by: Wechsung, F., Hartje, V., Kaden, S., Venohr, M., Hansjürgens, B., and Gräfe, P., Weißensee Verlag Berlin, 2012.
- Overland, J. E. and Preisendorfer, R. W.: A significance test for principal components applied to a cyclone climatology, *Mon. Weather Rev.*, 110, 1–4, doi:10.1175/1520-0493(1982)110<0001:ASTFPC>2.0.CO;2, 1982.
- Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., and Scipal, K.: Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, *Hydrol. Earth Syst. Sci.*, 10, 353–368, doi:10.5194/hess-10-353-2006, 2006.
- Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A., Gaume, E., Hlavcová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J., and Blöschl, G.: Seasonal characteristics of flood regimes across the Alpine-Carpathian range, *J. Hydrol.*, 394, 78–89, 2010.
- Perry, M. A. and Niemann, J. D.: Analysis and estimation of soil moisture at the catchment scale using EOFs, *J. Hydrol.*, 334, 388–404, 2007.
- Petrow, T., Zimmer, J., and Merz, B.: Changes in the flood hazard in Germany through changing frequency and persistence of circulation patterns, *Nat. Hazards Earth Syst. Sci.*, 9, 1409–1423, doi:10.5194/nhess-9-1409-2009, 2009.
- Philipp, A., Bartholy, J., Beck, C., Ercicum, M., Esteban, P., Fettweis, X., Huth, R., James, P., Jourdain, S., Kreienkamp, F., Krennert, T., Lykoudis, S., Michalides, S. C., Pianko-Kluczynska, K., Post, P., Álvarez, D. R., Schiemann, R., Spekat, A., and Tymvios, F. S.: Cost733cat – A database of weather and circulation type classifications, *Phys. Chem. Earth*, 35, 360–373, 2010.
- Preisendorfer, R. W.: *Principal component analysis in meteorology and oceanography*, *Developments in Atmospheric Science*, Vol. 17, edited by: Mobley, C. D., Elsevier, Amsterdam, 425 pp., 1988.
- Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, *Mon. Weather Rev.*, 100, 81–92, 1972.
- Prudhomme, C. and Geneviev, M.: Can atmospheric circulation be linked to flooding in Europe?, *Hydrol. Process.*, 25, 1180–1990, 2011.
- Ritchie, J. T.: Model for predicting evaporation from a row crop with incomplete cover, *Water Resour. Res.*, 8, 1204–1213, 1972.
- Rodda, H. J. E.: The development and application of a flood risk model for the Czech Republic, *Nat. Hazards*, 36, 207–220, 2005.
- Schirpke, H., Richter, I., and Rieß, B.: Analyse der Hochwassersituation im Oktober 1974/Dezember 1974/Januar 1975 in den Flußgebieten der Oberen Elbe, der Schwarzen Elster und den Mulden, in: *Ablauf der Winterhochwässer 1974/75 im Gebiet der DDR*, *Mitteilungen des Instituts für Wasserwirtschaft*, 41, 33–119, 1978.
- Sivapalan, M., Blöschl, G., Merz, R., and Gutknecht, D.: Linking flood frequency to long-term water balance: Incorporating effects of seasonality, *Water Resour. Res.*, 41, W06012, doi:10.1029/2004wr003439, 2005.
- Uhlemann, S., Thieken, A. H., and Merz, B.: A consistent set of trans-basin floods in Germany between 1952–2002, *Hydrol. Earth Syst. Sci.*, 14, 1277–1295, doi:10.5194/hess-14-1277-2010, 2010.
- Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 1 – Rainfall periods and flood development, *Weather*, 58, 371–377, 2003a.
- Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 2 - Synoptic causes and considerations with respect to climate change, *Weather*, 58, 434–442, 2003b.
- USDA Soil Conservation Service: *National Engineering Handbook*, 1972.
- Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, *Remote Sens. Environ.*, 70, 191–207, 1999.
- Ward Jr., J. H.: Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, 58, 236–244, 1963.
- Witrock, V. and Ripley, E. A.: The predictability of autumn soil moisture levels on the Canadian Prairies, *Int. J. Climatol.*, 19, 271–289, 1999.