Hydrol. Earth Syst. Sci. Discuss., 9, 10053–10094, 2012
www.hydrol-earth-syst-sci-discuss.net/9/10053/2012/
doi:10.5194/hessd-9-10053-2012

Hydrology and Earth System Sciences Discussions

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

Received: 22 August 2012 – Accepted: 22 August 2012 – Published: 5 September 2012

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

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 to classify soil moisture as a key variable of pre-event catchment conditions and to investigate 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 the spatial heterogeneous antecedent catchment conditions. By clustering the leading PCs, we detected 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. The results underline the importance of large-scale pre-event catchment conditions in flood initiation.

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 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

HESSD
9, 10053–10094, 2012

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), 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) what 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 ($< 1\,km^2$) 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, 1999). Moreover, no studies are currently 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

HESSD
9, 10053–10094, 2012

conditions, it is examined 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 (Brocca et al., 2009)), 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 (API) (Merz et al., 2006; Brocca et al., 2009), Gradex method (Merz and Blöschl, 2008), 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. It is assumed that the implemented rainfall-runoff model incorporates the hydrological processes that enable the estimation of soil moisture. Afterwards, a principal components 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). For a physical interpretation, PCs have been related to topography, land use and soil properties (Jawson and Niemann, 2007; Kim and Barros, 2002; Ibrahim and Huggins, 2011), land management (Korres et al., 2010) as well as teleconnection indices such as the Southern Oscillation or the North Atlantic Oscillation Index (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

HESSD
9, 10053–10094, 2012

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 km^2 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 $m^3 s^{-1}$ of river discharge (60 %) at its confluence with the Elbe river, the Eger/Ohře (38 $m^3 s^{-1}$), the Mulde (67 $m^3 s^{-1}$), the Saale (117 $m^3 s^{-1}$), the Schwarze Elster (21 $m^3 s^{-1}$) and the Havel (114 $m^3 s^{-1}$). The climate is temperate. The Elbe river basin is situated in a transition zone between maritime (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 winter time, 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

HESSD
9, 10053–10094, 2012

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 were 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

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 as well as the physical interpretation of the PCs.

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

HESSD
9, 10053–10094, 2012

available in the Czech part. Catchment size varies between 104 km^2 and 131 950 km^2 . 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 were used for the calibration of the rainfall-runoff model (Fig. 1, red dots).

3 Methodology

First, the soil profile water content is identified with a semi distributed rainfall-runoff model. PCA is applied to reduce the data dimensionality to specific daily spatial soil moisture patterns expressing large part of the spatial variability of the soil moisture dynamics (PCs). Correlation analysis provides a physical interpretation of the identified PCs. Cluster analysis detects 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. The analysis time period is 1 November 1951 to 31 October 2003.

3.1 Hydrological modeling

3.1.1 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 on the subbasin centroids.

HESSD
9, 10053–10094, 2012

The snow routine and the soil water balance are calculated on the hydrotops and river routing on the basin scale. 1945 subbasins with a median catchment size of 33 km^2 (minimum 2 km^2, maximum 1034 km^2) 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. The percolation from the bottom soil layer to the shallow aquifer is considered as recharge. 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 and soil transpiration are calculated separately as a function of the leaf area index according to Ritchie (1972). The actual soil transpiration is estimated on the upper 0.3 m of the soil zone. As long as the accumulated soil transpiration is below 6 mm, actual soil transpiration equals potential soil transpiration. If the accumulated soil transpiration is above the threshold value, potential soil transpiration is reduced in dependence of the number of days the threshold value is exceeded. In the case of a snow cover, the soil transpiration 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.

HESSD
9, 10053–10094, 2012

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). Surface runoff and the sum of subsurface and groundwater are routed separately.

3.1.2 Model calibration

The simulated spatio-temporal soil moisture patterns depend on model parameterization and several parameter combinations can lead to the same model performance (equifinality) (Beven and Binley, 1992). Hence, a Monte Carlo uncertainty analysis is carried out to identify an ensemble of parameter sets that lead to behavioral model performances. Nine sensitive parameters controlling snow accumulation and melt, potential evapotranspiration, saturated hydraulic conductivity, recharge as well as discharge routing are calibrated over the period 1981–1989.

For model calibration, the Elbe river basin is subdivided into 27 regions (three located in the Czech Republic), assuming homogeneous parameterization within each region (Fig. 1, red dots). 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 for model calibration:

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

HESSD
9, 10053–10094, 2012

where q_c and q_o are simulated and observed discharge, respectively. $\bar{q}_o$ is the average observed daily discharge in the calibration 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.

3.2 Principal component analysis

PCA enables to identify spatial patterns of soil moisture variability by reducing the dimensionality of a data set with numerous interrelated variables while maintaining most of the data variability by a new set of uncorrelated variables, the principal components (PCs) (Joliffe, 1986).

Daily soil moisture values of the soil profile **X** for each subbasin were simulated by the rainfall-runoff model resulting in i data matrices **X** ($m \times n$), where m is the number of observations in time (i.e. 18 993), n is the number of subbasins (i.e. 1945 till gauge Wittenberge) and i refers to the number of behavioral parameter combinations (Monte Carlo parameter sets).

$$\mathbf{X}_i = \begin{matrix} \mathbf{X}_{1,1} \cdots \mathbf{X}_{1,n} \\ \vdots \quad \ldots \quad \vdots \\ \mathbf{X}_{m,1} \cdots \mathbf{X}_{m,n} \end{matrix} \tag{2}$$

For the comparability of the profile soil moisture values of different soil types, the values are standardized by the field capacity of each soil type. In the following, the standardized profile soil moisture is termed soil moisture index (SMI).

To include the parameter uncertainty of the rainfall-runoff model in the PCA, the soil moisture simulations are arranged consecutively in time reshaping **X** into a matrix $\mathbf{X}^*$ of size $((i \times m) \times n)$. PCA is performed on the spatial linear Pearson correlation matrix $\mathbf{R}(n \times n)$ of $\mathbf{X}^*$. All locations are equally weighed since they have unit variance. As **R** is square and symmetric and therefore diagonalizable, one can identify the eigenvectors

HESSD
9, 10053–10094, 2012

u (weights applied to the original data **X**) and the eigenvalues diag ($\boldsymbol{\lambda}$) of the matrix **R**:

$$\mathbf{Ru} = \boldsymbol{\lambda}\mathbf{u} \tag{3}$$

by solving

$$(\mathbf{R} - \boldsymbol{\lambda}\mathbf{I})\mathbf{u} = 0 \tag{4}$$

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

$$\mathrm{ev}_k = \frac{\lambda_k}{\sum_{k=1}^{n} \lambda_k} \tag{5}$$

Hence, the leading first eigenvector points in the direction of the highest variance of $\mathbf{X}^*$ and the next eigenvector explains the subsequent highest variance with the condition being orthogonal to the already identified eigenvector.

The principal components PC are obtained by projecting the standardized (zero mean, unit variance) data matrix of $\mathbf{X}^*$ onto the eigenvector $\boldsymbol{u}_k$.

$$\mathrm{PC}_k = \mathbf{X}^*_{\mathrm{std}} \boldsymbol{u}_k \tag{6}$$

As $\mathbf{X}^*$ compromises all behavioral Monte Carlo runs, uncertainty bounds of each PC_k are obtained by decomposing PC_k of size (($i \times m$) × 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.

HESSD
9, 10053–10094, 2012

For further details on PCA see e.g. Hannachi et al. (2007), Joliffe (1986) or Preisendorfer (1988).

Additionally, we studied the relationship between the different principal components PC_k and the daily climate variables precipitation, air temperature and antecedent precipitation index (API). The API is a weighted sum of precipitation in the previous / days. The weights are a reciprocal function of time, with the most recent day receiving the largest weight.

$$API_l = \sum_{b=1}^{l} prec_{(l-b+1)} \times 0.9^b \; [l \in 3,5,7,21,31] \tag{7}$$

For all subbasins, the daily climate variables as well as the API are standardized to a mean value of zero and a standard deviation of one ($clim^*_{std}$) and projected onto the eigenvector $\boldsymbol{u}_k$ of the SMI.

$$clim_{PC\,k} = clim^*_{std} \boldsymbol{u}_k \tag{8}$$

$clim_{PC\,k}$ is then correlated with the median value of the decomposed PCs of SMI. The non-parametric Spearman's rank correlation coefficient $\delta[\delta\varepsilon - 1...1]$ was used as the data are not necessarily normally distributed.

3.3 Cluster analysis

In a subsequent step, cluster analysis was implemented to identify days of similar soil moisture patterns. Up to 15 of the leading PCs were used as variables in the cluster analysis to evaluate the influence of explained variance on the clustering. The weight assigned to a particular PC depends on its explained variance, i.e. the first PC has the highest weight.

The applied distance metric is the Euclidian distance to give more weight to the high differences. In a subsequent step, the hierarchical Ward cluster algorithm merges

HESSD
9, 10053–10094, 2012

those two clusters (i.e. days D) that offer the smallest increase in variance leading to homogenous clusters of similar soil moisture patterns. As the merging continues until all days are in a single cluster, cluster validation techniques are implemented to determine the optimum number of clusters. For an overview of the different cluster validation methods (stopping rules) see e.g. Milligan and Cooper (1985). We applied the following six measures: Calinski and Harabasz (1974), Hartigan (1975), Davies and Bouldin (1979), Krzanowski and Lai (1988), Average Silhouette (Kaufman and Rousseeuw, 1990) and Marriott (1971).

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 was not feasible to accomplish a cluster analysis directly on the leading PCs. Thus, we executed the cluster analysis in two steps. First, the leading PCs were decomposed for each behavioral Monte Carlo run. Afterwards, 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:

$$p_D = \max\left(\frac{|D \in t|}{i}\right) \tag{9}$$

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

$$p_t = \text{median}(p_D \in t) \tag{10}$$

HESSD
9, 10053–10094, 2012

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.

3.4 Flood event identification

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. To investigate 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). 114 discharge time series were used for flood identification (Fig. 1, dots). A flood event occurs 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 timeframe 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 at least four days are between the last occurrence of a significant peak of the previous flood and the first significant peak of the following flood event. 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).

HESSD
9, 10053–10094, 2012

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.

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 rainfall-runoff model was calibrated against observed discharge at 27 gauges over the 9-yr period 1981–1989. Validation was carried out at 26 gauges over 1951–1980 and 1990–2003, respectively.

The application of the Monte Carlo approach resulted in 38 behavioral Monte Carlo 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

HESSD
9, 10053–10094, 2012

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).

4.3 Soil moisture pattern classification

4.3.1 Principal component analysis

The simulated profile soil moisture of the behavioral parameter sets was utilized to identify soil moisture pattern types using PCA. The leading 43 out of 1945 PCs are significantly different from a random field. These significant PCs explain 97.66 % of the total variance. Figure 3 shows the explained variance of the leading 20 PCs. The leading five PCs and their corresponding eigenvectors are displayed in Fig. 4. The eigenvectors indicate geographic regions of simultaneous anomalies in the soil moisture index (SMI). The minimum and maximum values of the PCs correspond to the parameter uncertainty introduced by the rainfall-runoff model.

The first PC which explains 75.71 % of the total variance (Fig. 3) shows the lowest spatial variability across the catchment and has a seasonal behavior (Fig. 4, top). The influence of the parameter uncertainty on the PC is small. The subsequent PCs have spatially heterogeneous eigenvectors as well as a damped and lagged seasonal behavior (Fig. 4). The second PC explains 8.60 % of the total variance and shows 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 shows negative loadings. In the temporal progression seasonality is still

HESSD
9, 10053–10094, 2012

visible. Compared to the first PC, the influence of the parameter uncertainty on the PC increased although the general behavior is the same. The third PC has high positive loadings in the Saale region and the mountainous area of the catchment. There is no apparent periodic behavior in the PC. The PC explains 1.87 % of the total variability. The fourth PC has positive anomalies in large parts of Saale region, small positive anomalies in the central Czech part of the catchment and high negative anomalies in the downstream Havel region. The explained variance is 1.49 %. The influence of the parameter uncertainty on the PC is the highest of the five presented PCs although the different parameter sets agree in their general behavior. The fifth PC shows negative anomalies in the Czech mountains and parts of the Saale region. Positive anomalies are found in the centre of the Elbe catchment. The explained variance decreases to 1.33 %. In total, the presented PCs explain 89 % of the entire variance.

The relationship between the different PCs and daily climate variables is studied by means of the Spearman's rank correlation coefficient δ. Figure 5 summarizes the most important findings exemplified for the leading five PCs. Apparently, there is a difference between the first and the remaining PCs in the Spearman's rank correlation coefficient δ. On the one hand, the first PC is negatively correlated with mean daily air temperature ($\delta = -0.76$). On the other hand, daily precipitation as well as the weighted sums of precipitation (API) are uncorrelated with the first PC. The remaining PCs are uncorrelated with mean daily air temperature ($\delta < 0.09$) and the correlation strength of precipitation increases the more days are aggregated. The leading sixth to 20th PCs (not shown) behave similar in their correlation sign and strength.

4.3.2 Cluster analysis

The PCA reduced the soil moisture data to characteristic soil moisture pattern types (PCs). 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.

HESSD
9, 10053–10094, 2012

Six cluster validation methods were implemented to identify the optimum cluster number. However, due to diverging results between the different validation methods, no optimum number of clusters could be identified. Therefore, the number of clusters remained arbitrary in terms of cluster validation methods. Additionally, the probability of cluster membership p_t (Eq. 10) was used to evaluate the influence of model parameterization on the different PC-cluster combinations. Figure 6 (left) displays the median p_t of different PC-cluster combinations. A small cluster number leads in the median to well distinguishable clusters independent of model parameterization (high median p_t). The median p_t decreases up to a division in approximately 12 clusters. For a higher number of clusters, the median p_t remains stable. In general, the median p_t is independent of the number of PCs in the 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, we restrict the analysis to the clustering of the leading four PCs since the median cluster probability is stable for a higher number of clusters and already 86.67 % of explained variance is included in the clustering. For different number of clusters, clusters with a very high p_t and thus independent of model parameterization and clusters with low p_t (< 0.5) and thus dependent on model parameterization can be found in almost all cases (Fig. 6, right). In a subsequent step, we restrict the analysis to ten clusters since the median p_t peaks for ten clusters (0.74) and decays for more clusters.

The soil moisture index (SMI) patterns of the cluster centroids when separating the leading four PCs into ten clusters are presented in Fig. 7. Cluster one, seven and eight show the driest soil moisture patterns. Between 87.8 % and 97.5 % of the catchment area has a SMI below 0.6. Clusters two and ten have intermediate soil moisture patterns. Around 72 % of the catchment area has a SMI below 0.6. High SMI values are mainly restricted to the mountainous areas. Around 4 % of the catchment area has a SMI above 0.8. The remaining clusters change consecutively into highly saturated soil moisture patterns. In general, saturation increases from the sixth to the fifth, to the

HESSD
9, 10053–10094, 2012

fourth, to the third and finally to the ninth soil moisture pattern. Initially, a high SMI is restricted to the mountainous areas and intermediate values 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 a SMI above 0.8. Finally, cluster nine shows high SMI over the entire basin. 43.8 % of the catchment area has a SMI above 0.8 and 90.1 % of the catchment area has a SMI above 0.6.

For a better understanding of the PCs' influence on each clusters spatial SMI distribution, Fig. 8 displays the PC distributions of the first and second PC according to the different clusters. In general, the first PC separates the patterns into clusters of high SMI i.e. cluster three, four, five as well as nine and clusters of low SMI i.e. cluster one, two, seven, eight as well as ten. The second PC separates the patterns according to the spatial SMI distribution in the catchment. Cluster three, five, eight and ten have a north-south partition in the SMI, whereas cluster one, two, four, seven and nine have approximately uniformly distributed SMI in the catchment. Due to the small amount of explained variance by the subsequent PCs, they show only minor differences between the different clusters and it is not possible to attribute these small differences to pattern characteristics.

Table 1 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 the entire days. Clusters two and seven are the smallest clusters containing around 5 % of the entire 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

HESSD
9, 10053–10094, 2012

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. 9 (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, cluster 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

Hitherto, the classification of the large-scale soil moisture conditions didn't take the flood initiation of the different soil moisture pattern types into account. 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 of the highest percentage (52.9 %) of flood start days (Table 1). Clusters seven, eight and ten contain less than 2 % of flood start days each. 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. Approximately, every 100th day in cluster nine initiates a flood. In comparison, 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. Cluster seven contains nearly no flood start days.

The relative frequency of flood start days within a respective month and pattern type are presented in Fig. 9 (bottom). In summer, the highest relative frequencies of flood start days can be found and a multiplicity of pattern types is related to 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

HESSD
9, 10053–10094, 2012

frequency of flood start days is approximately constant from January until April/May. Although the two clusters have approximately the same seasonal distribution (Fig. 9, top), cluster nine is primarily relevant for flood initiation in winter whereas cluster three is of primary importance in summer time flood initiation. In June, cluster three has the highest relative frequency of flood start days of all clusters. This frequency shift between the monthly occurrence of cluster days in general (Fig. 9, top) and flood start days inside a respective cluster (Fig. 9, bottom) also appears in case of cluster 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 winter time, 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 transforming into wetter patterns after the flood start date.

Finally, the relationship between the soil moisture patterns types and flood severity, a combined measure of flood magnitude and extent, is analyzed (Table 1). 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.

HESSD
9, 10053–10094, 2012

5 Discussion

An automated classification of regional soil moisture patterns enabled a probabilistic linkage between flood occurrence and large-scale soil moisture patterns as proxy for the catchment state in the Elbe river basin.

5.1 Soil moisture pattern classification

PCA identified subbasins having the same temporal soil moisture anomalies. The large amount of total variability explained by the first PC (Fig. 3) 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 may be attributable to seasonal changes in evapotranspiration leading to soil moisture depletion in summer and rise in winter and spring (Parajka et al., 2010). This is also reflected in the high negative correlation of the first PC with daily air temperature (Fig. 5). Additionally, snowmelt may have an impact. The spatial weights (eigenvectors) applied to the original soil moisture data are positive all over the catchment (Fig. 4, top left). Thus, all subbasins respond in the same direction in terms of seasonality indicating considerable similarity in the processes controlling soil moisture variability. Heterogeneous hydrological processes and catchment properties are overruled by seasonality. Perry and Niemann (2007) reported similar findings for the leading PC. Moreover, the leading PC is uncorrelated with precipitation (Fig. 5). 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 contrast, the subsequent PCs are a measure of the disparity describing local variations departing from the regional value. In case of the subsequent PCs, no relationship with air temperature is detected whereas accumulated precipitation shows a small but continuously increasing correlation, the more days are summed. Therefore,

HESSD
9, 10053–10094, 2012

the subsequent PCs are a measure of the catchment memory, i.e. the previous spatial distribution of precipitation.

To take into account the parameter uncertainty introduced by the rainfall-runoff model, PCA was applied on an ensemble of daily soil moisture patterns. 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 which were attributed to the spatial diverse catchment memory. This reveals a greater impact of model parameterization on spatial heterogeneous hydrological processes, as e.g. infiltration and soil storage, than on the seasonal progression of soil moisture.

As the rainfall-runoff 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 model calibration artifact. A comparison of the retrieved soil moisture patterns with auxiliary data such as those obtained using remote sensing would enable an additional verification. Nevertheless, there is a physical meaning behind the soil moisture patterns as soil and land use properties are included in the model. 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.

A cluster analysis identified days of similar soil moisture patterns, i.e. distinct soil moisture pattern types. Each soil moisture pattern type can be attributed to a pattern frequency, a pattern persistence and seasonal characteristics (Table 1, Fig. 9) in the same manner as in weather type classifications (e.g. Philipp et al., 2010). The dominant modes in the pattern classification are amount and spatial distribution (uniform vs. north-south) of the SMI (Fig. 8).

HESSD
9, 10053–10094, 2012

The weakness of the subjective choice of the number of PCs as well as of the number of clusters (Philipp et al., 2010) was affirmed by the diverging results of the cluster validation methods. 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. 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

The flood identification differed from the commonly applied approach that is based on identifying annual maximum or peak over threshold series at a single gauge. 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 due to the requirement of an observed 10-yr flood (POT). Furthermore, the flood event set may be biased towards flooding in the German part of the Elbe catchment as only three 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 case, i.e. limiting the event

HESSD
9, 10053–10094, 2012

identification to 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) didn't 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, the results are heterogeneous and reveal the different flood generating processes. On the one hand, the median flood severity increased for patterns of high SMI (Table 1). 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 Genevier (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 soil moisture pattern types and their attributed flood generating processes vary seasonally. In winter, flooding is related to high soil moisture content in the entire basin (Fig. 7, cluster nine). The respective soil moisture pattern is persistent before and after the flood start date. In summer, the flood initiating soil moisture patterns are more variable and less persistent reflecting the different flood generating processes. Flooding is either initiated by high soil moisture content over large parts of the basin (Fig. 7, cluster three) or by relatively dry catchment conditions (Fig. 7, cluster eight). Nevertheless, in addition to soil moisture other patterns, in particular precipitation, are relevant for flood initiation. This is indicated by the seasonal cluster distribution and the deviating seasonality of the flood start days inside the cluster (Fig. 9, pattern type three). Furthermore, extending the pattern classification approach to e.g. circulation patterns, the role of saturated soils limited to parts of the catchment when receiving rainfall may be pronounced.

HESSD
9, 10053–10094, 2012

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 the catchment state on floods and flood severity is expected to depend on various factors such as season as well as flood type (e.g. snowmelt flood, long-duration rainfall 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 to classify the hydrological catchment conditions and to 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 principal component analysis, the dimensionality of the soil moisture data was reduced to four PCs representing 86.67 % of the total variability. The first PC can be linked to the seasonally wetting and drying of the catchment, the successive PCs represent the spatial variation of antecedent catchment conditions. In a second step, cluster analysis assessed the similarity in daily soil moisture distribution within the catchment, and assigned each day of the investigation period 1951–2003 to one of ten clusters. In parallel, 94 start days of large-scale flood events were identified from a data set of 114 discharge gauges and the relevance of the soil moisture pattern types for flood events were derived. The ten clusters of daily soil moisture patterns are not equally associated with flood occurrence. For instance, cluster nine accumulates over 50 % of all flood start days and shows the highest median flood severity. It is the biggest cluster containing 24 % of all days, and is characterized by highest

HESSD
9, 10053–10094, 2012

persistence and highest soil moisture across the catchment. Similarly, cluster three is characterized by high soil moisture and is related to increased occurrence of flooding. However, although both clusters have similar seasonal distribution, cluster nine is primarily responsible for flood occurrence in winter and cluster three for flood occurrence in summer. Other clusters, e.g. cluster seven, are hardly linked to large-scale floods. Hence, the relationship between large-scale soil moisture patterns and floods varies clearly between clusters and during the year. In winter, large-scale floods in the Elbe catchment are initiated by overall high soil moisture content, whereas in summer the flood initiating soil moisture patterns are diverse, less stable in time and strongly dependent on parameterization of the hydrological model.

These results underline the importance of catchment state for flood initiation and severity. They show that large-scale soil moisture patterns can be identified which frequently occur before the onset of floods. The occurrence of these soil moisture patterns does not necessarily lead to floods, but the probability of occurrence of a large-scale flood is increased.

Future work will extend the pattern classification approach not only to circulation patterns but also to snow. This 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.

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, 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.

Calinski, R. B. and Harabasz, J.: A dendrite method for cluster analysis, Commun. Stat., 3, 1–27, 1974.

CORINE European Environment Agency: Land Cover Data Set, 2000.

Davies, D. L. and Bouldin, D. W.: A cluster separation measure, IEEE T. Pattern Anal., PAMI-1, 224–227, 1979.

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.

Hartigan, J. A.: Clustering Algorithms, John Wiley & Sons, Inc., New York, 1975.

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.

HESSD
9, 10053–10094, 2012

Hauptamt für Hydrologie: Witterung und Wasserstände im Juli 1954, Wasserwirtsch. Wassertech., 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.
Jacobeit, 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.
Joliffe, I. T.: Principal component analysis, Springer Series in Statistics, Springer, New York, 1986.
Kaufman, L. and Rousseeuw, P. J.: Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, New York, 1990.
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.
Krzanowski, W. J. and Lai, Y. T.: A criterion for determining the number of groups in a data set using sum-of-squares clustering, Biometrics, 44, 23–34, 1988.
Maidment, D. R.: Handbook of Hydrology, McGraw-Hill, Inc., New York, 1993.

HESSD
9, 10053–10094, 2012

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.
Marriott, F. H. C.: Practical problems in a method of cluster analysis, Biometrics, 27, 501–514, 1971.
Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39, 1340, 2003.
Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, Water Resour. Res., 44, W08432, 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.
Milligan, G. W. and Cooper, M. C.: An examination of procedures for determining the number of clusters in a data set, Psychometrika, 50, 159–179, 1985.
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, 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, Berlin, 2012.
Overland, J. E. and Preisendorfer, R. W.: A significance test for principal components applied to a cyclone climatology, Mon. Weather Rev., 110, 1982.
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.

HESSD
9, 10053–10094, 2012

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. Hazard Earth Sys., 9, 1409–1423, 2009.
Philipp, A., Bartholy, J., Beck, C., Erpicum, 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, 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 Genevier, 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, Mitt. Inst. Wasserwirtsch., 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, 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.

HESSD
9, 10053–10094, 2012

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.

Wittrock, V. and Ripley, E. A.: The predictability of autumn soil moisture levels on the Canadian Prairies, Int. J. Climatol., 19, 271–289, 1999.

HESSD

9, 10053–10094, 2012

Table 1. 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 (172)	4.8 (172)	11.0 (393)	8.2 (292)	6.7 (241)	8.2 (293)	0.2 (7)	1.9 (69)	52.9 (1888)	1.3 (45)	100 (38 x 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

HESSD
9, 10053–10094, 2012

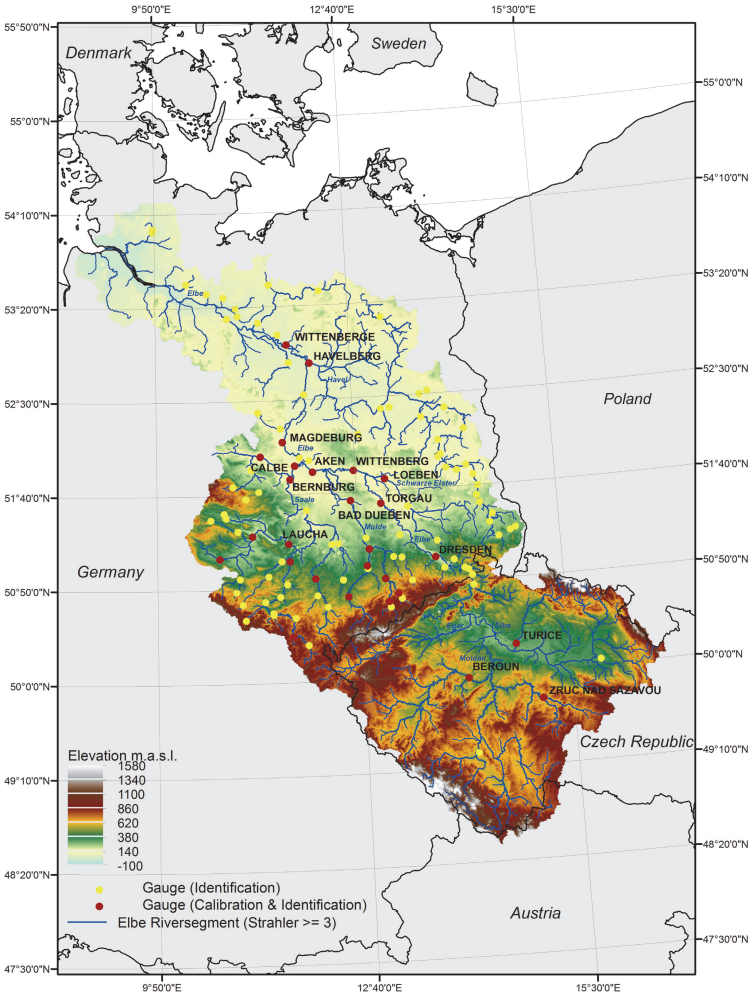

Fig. 1. Elbe river basin.

HESSD
9, 10053–10094, 2012

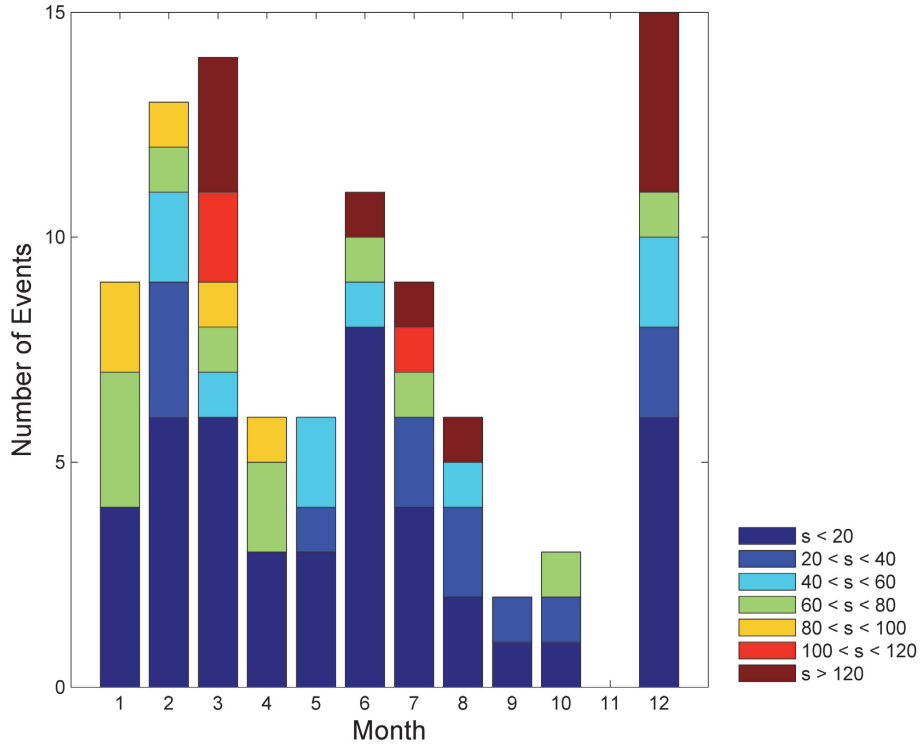

Fig. 2. Number of flood start dates per month and severity class *s*.

HESSD
9, 10053–10094, 2012

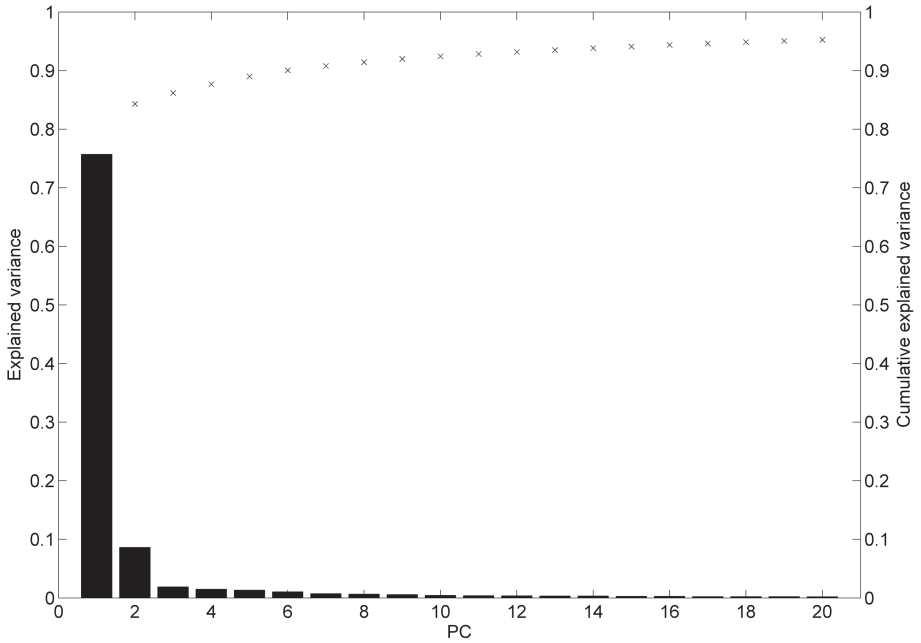

Fig. 3. Explained variance (left axes) and cumulative explained variance (right axes) of leading PCs.

HESSD
9, 10053–10094, 2012

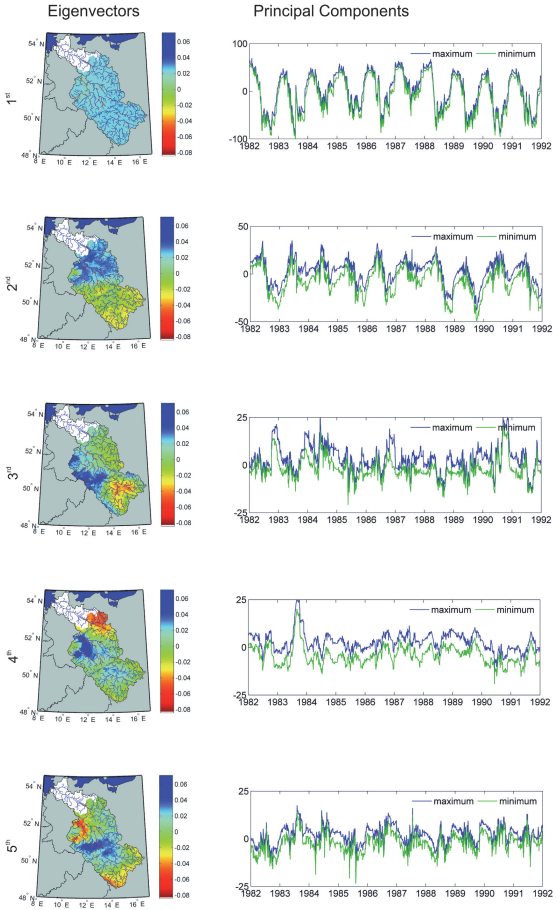

Fig. 4. Eigenvectors of the leading five PCs (left). PCs for the sub-period 1982–1991 (right). Minimum and maximum values correspond to the parameter uncertainty introduced by the rainfall-runoff model.

HESSD
9, 10053–10094, 2012

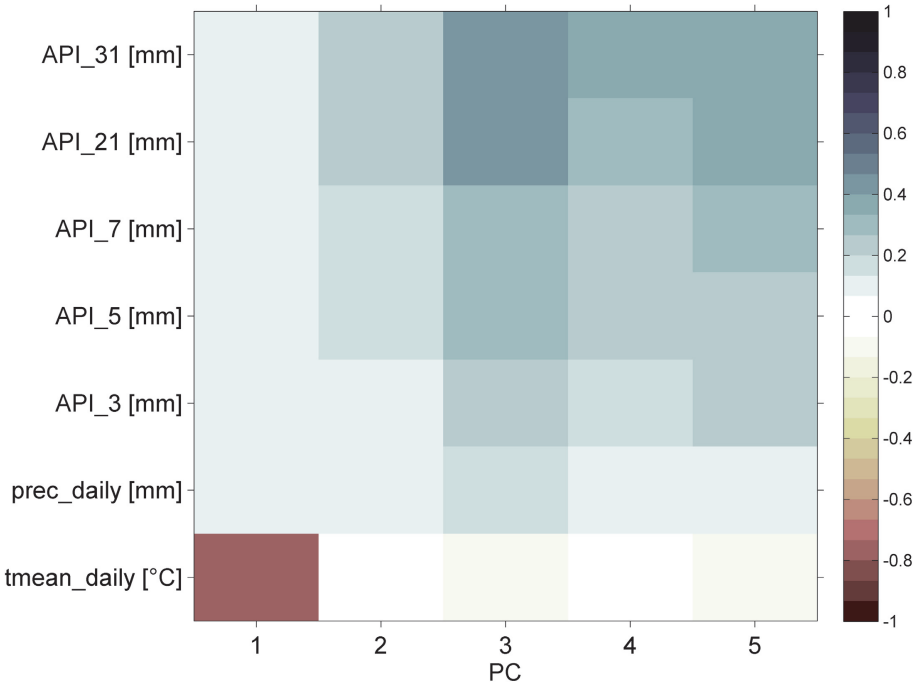

Fig. 5. Spearman's rank correlation coefficient δ correlating PCs and temporal catchment attributes.

HESSD
9, 10053–10094, 2012

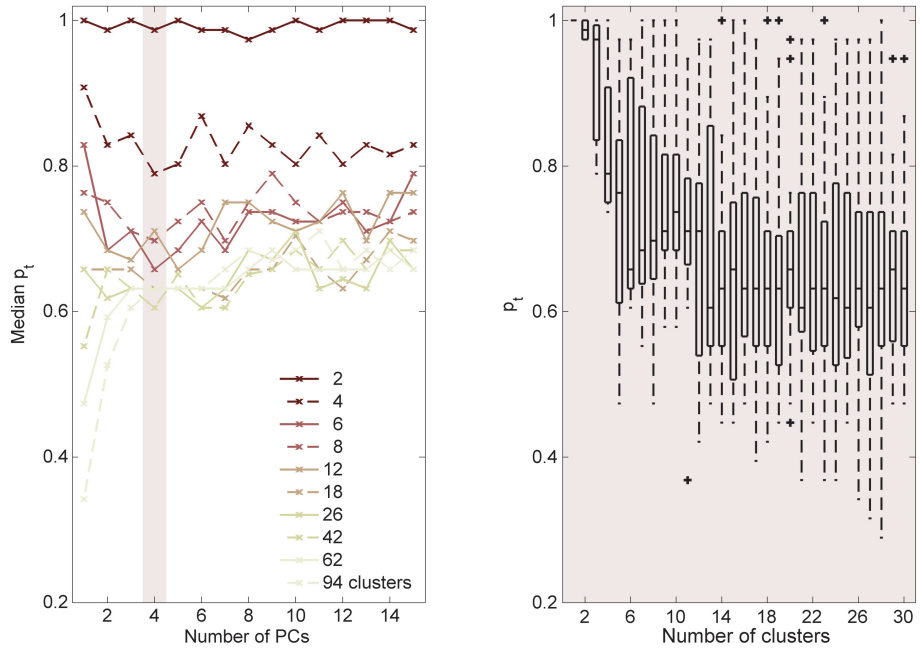

Fig. 6. Median probability of cluster membership p_t of different PC-cluster combinations (left). Distribution of p_t for different numbers of clusters when clustering the leading four PCs (right).

HESSD
9, 10053–10094, 2012

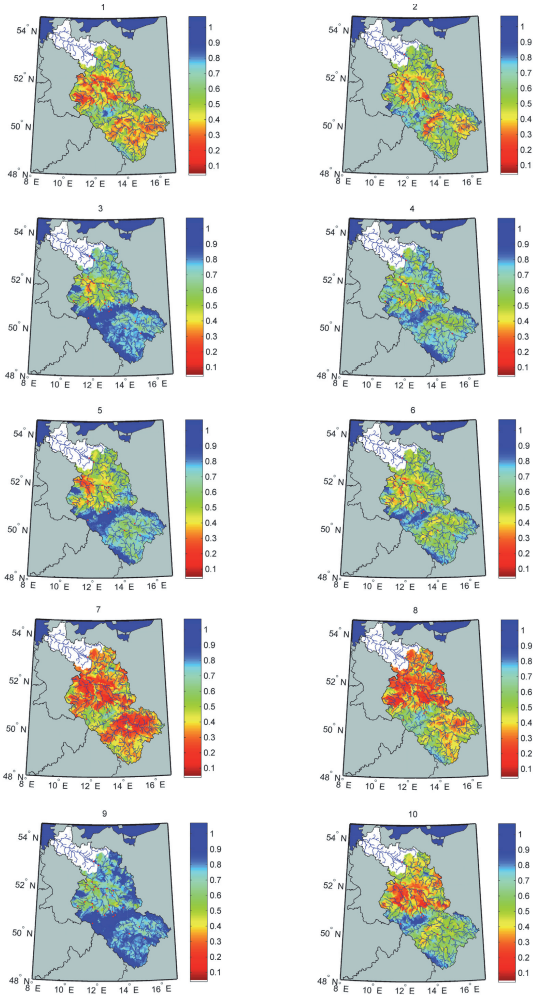

Fig. 7. Soil moisture index (SMI) patterns of cluster centroids.

HESSD
9, 10053–10094, 2012

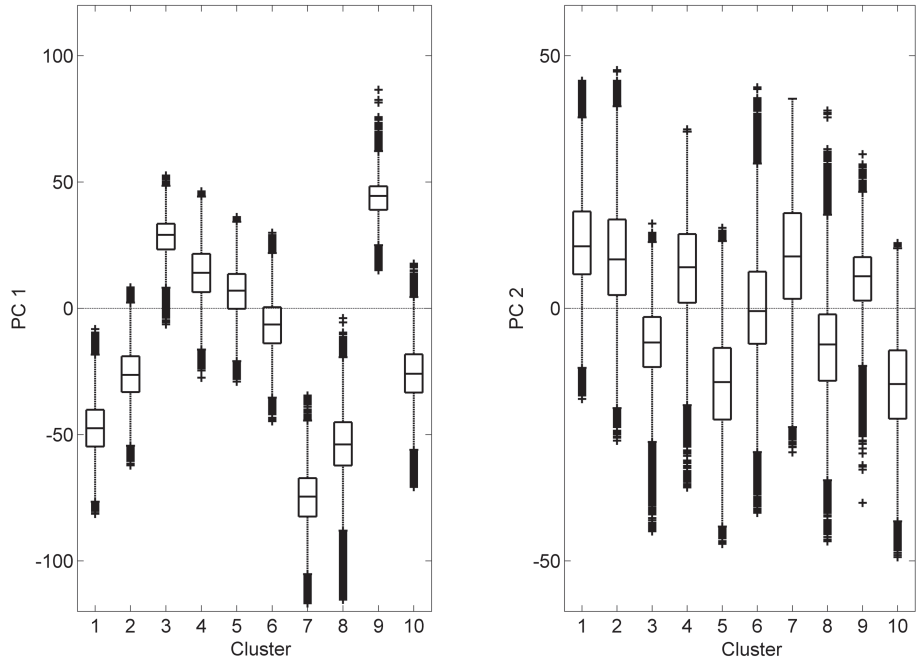

Fig. 8. Distribution of first (left) and second (right) PC per cluster.

HESSD
9, 10053–10094, 2012

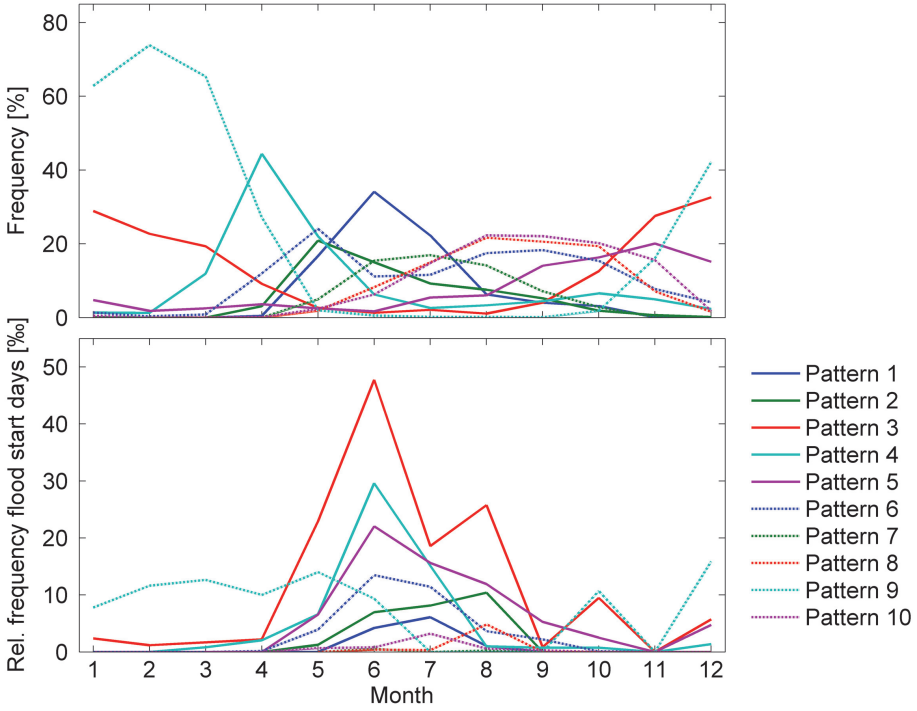

Fig. 9. Frequency of occurrence per month of different soil moisture pattern types [%] (top). Relative frequency of flood start days per month and respective soil moisture pattern type [‰] (bottom).

HESSD
9, 10053–10094, 2012