UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Interannual drought index variations in Central Europe related to the large-scale atmospheric circulation - application and evaluation of statistical downscaling approaches based on circulation type classifications

Christoph Beck, Andreas Philipp, Jucundus Jacobeit

Angaben zur Veröffentlichung / Publication details:

Beck, Christoph, Andreas Philipp, and Jucundus Jacobeit. 2015. "Interannual drought index variations in Central Europe related to the large-scale atmospheric circulation - application and evaluation of statistical downscaling approaches based on circulation type classifications." *Theoretical and Applied Climatology* 121 (3-4): 713–32. https://doi.org/10.1007/s00704-014-1267-z.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

Interannual drought index variations in Central Europe related to the large-scale atmospheric circulation—application and evaluation of statistical downscaling approaches based on circulation type classifications

Christoph Beck · Andreas Philipp · Jucundus Jacobeit

Abstract This contribution investigates the relationship between the large-scale atmospheric circulation and interannual variations of the standardized precipitation index (SPI) in Central Europe. To this end, circulation types (CT) have been derived from a variety of circulation type classifications (CTC) applied to daily sea level pressure (SLP) data and mean circulation indices of vorticity (*V*), zonality (*Z*) and meridionality (*M*) have been calculated. Occurrence frequencies of CTs and circulation indices have been utilized as predictors within multiple regression models (MRM) for the estimation of gridded 3-month SPI values over Central Europe, for the period 1950 to 2010. CTC-based MRMs used in the analyses comprise variants concerning the basic method for CT classification, the number of CTs, the size and location of the spatial domain used for CTCs and the exclusive use of CT frequencies or the combined use of CT frequencies and mean circulation indices as predictors. Adequate MRM predictor combinations have been identified by applying stepwise multiple regression analyses within a resampling framework. The performance (robustness) of the resulting MRMs has been quantified based on a leave-one-out cross-validation procedure applying several skill scores. Furthermore, the relative importance of individual predictors has been estimated for each MRM. From these analyses, it can be stated that model skill is improved by (i) the consideration of vorticity characteristics within CTCs, (ii) a relatively small size of the spatial domain to which CTCs are applied and (iii) the inclusion of mean circulation indices. However, model skill exhibits distinct variations between seasons and regions. Whereas promising skill can be stated for the western and northwestern parts of the Central European domain, only unsatisfactory skill is reached in the more continental regions and particularly during summer. Thus, it can be concluded that the presented approaches feature the potential for the downscaling of Central European drought index variations from the large-scale circulation, at least for some regions. Further improvements of CTC-based approaches may be expected from the optimization of CTCs for explaining the SPI, e.g. via the inclusion of additional variables in the classification procedure.

C. Beck (✉) · A. Philipp · J. Jacobeit
Institute of Geography, University of Augsburg, Alter Postweg 118, D-86135 Augsburg, Germany
e-mail: christoph.beck@geo.uni-augsburg.de

1 Introduction

Droughts are an important feature of Central European climate, having wide-ranging significant impacts on natural and socio economic systems. With respect to the drought event in 2003 affecting large parts of Europe, for example Fink et al. (2004) documented the adverse effects on agriculture, forestry and energy production—just to name a few highly relevant points.

A general definition of drought in a meteorological sense is the deficiency in precipitation compared to long-term average conditions (Hayes et al. 2011). However, depending on which aspects of droughts are primarily considered, varying drought definitions incorporating different meteorological, agricultural, hydrological or socioeconomic parameters have been developed (e.g. Heim 2000, Heim 2002, Keyantash and Dracup 2002). Focusing on the meteorological component of droughts, from the varying precipitation based indices that have been defined for the quantification of drought events (e.g. Keyantash and Dracup 2002), the standardized precipitation index SPI (McKee et al. 1993, McKee et al. 1995) has preferably been used in recent years in numerous studies. Most recently, the utilization of the SPI has been further

fostered by the decision of the WMO to recommend the SPI for characterising meteorological drought by national meteorological and hydrological services around the world (Hayes et al. 2011). The spatiotemporal variability and the spatial coherence of droughts in Europe has been analysed, based on station or gridded SPI data, for example by Lloyd-Hughes and Saunders (2002), Bordi et al. (2007), Bordi et al. (2009), Santos et al. (2010), Hannaford et al. (2011) and Lloyd-Hughes and Saunders (2002). Applications of the SPI to the output of general circulation models (GCMs) and regional climate models (RCMs) have been presented for example by Loukas et al. (2007), Sienz et al. (2007), Burke and Brown (2008), Vasiliades et al. (2009), Bonsal et al. (2012), Heinrich and Gobiet (2012) and Sienz et al. (2012).

The SPI is solely based on precipitation data. Drought variability as captured by the SPI is therefore closely linked to large-scale atmospheric circulation features that essentially govern precipitation. Analysing and quantifying the relationships between SPI and atmospheric circulation is essential to improve the understanding of the development of drought events and their interannual and longer-term variations in frequency and intensity. Furthermore, once reliable quantitative assessments of the linkage between circulation and drought are acquired, this offers the opportunity to derive statistical transfer functions for downscaling of potential future drought dynamics on the basis of GCM scenarios.

Varying approaches have been used to investigate the relationships between atmospheric circulation variability and droughts in general and the SPI in particular. Based on the SPI, Bordi et al. (2007) characterised atmospheric circulation conditions during extremely dry/wet periods in Sicily via the calculation of respective anomaly maps for the 500 hPa geopotential height field. Bothe et al. (2010) determined characteristic patterns of geopotential height and zonal wind for dryness/wetness on the Tibetan plateau by regression analysis. Lopez-Moreno and Vicente-Serrano (2008) analysed the effect of positive and negative phases of the North-Atlantic Oscillation (NAO) on droughts in Europe. Vicente-Serrano and Lopez-Moreno (2006) and Raziei et al. (2012) related occurrence frequencies of circulation types (CTs), resulting from different circulation type classifications (CTCs), to SPI variations in Spain and Iran respectively. Both studies document a distinct relationship between variations in frequencies of CTs and SPI. In addition, CTCs as tool for describing and analysing the atmospheric circulation (e.g. Huth et al. 2008, Philipp et al. 2010) have been used in several other—non-SPI-based—studies for investigating the influence of large-scale circulation on drought. Stahl and Demuth (1999) pointed out the relevance of subjectively derived Grosswetterlagen (Hess and Brezowsky 1952, Gerstengarbe et al. 1999) for streamflow droughts in southern Germany. Using likewise the Grosswetterlagen classification, Trnka et al. (2009) analysed the influence of CTs on soil moisture in the Czech Republic and Pongracz et al. (2003) investigated the linkage between variations in CT frequencies and time series of the Palmer Drought Severity Index (PDSI) in Hungary. Fleig et al. (2010) compared several CTCs concerning their ability to capture hydrological drought events in northwestern Europe and Fleig et al. (2011) linked an objective version of the Hess-Brezowsky Grosswetterlagen (James 2007) to regional hydrological drought indices in northwestern Europe. An objective version of the Lamb weather types (Lamb 1972, Jenkinson and Collison 1977) was applied by Garcia-Herrera et al. (2007) to characterise atmospheric circulation conditions during a prolonged drought episode in the Iberian peninsula. Tran et al. (2002) developed a CTC utilizing self organizing maps (SOM) and related the resulting CTs to the occurrence of wet and dry conditions in Bulgaria. Pongracz et al. (1999) used principal component analysis with subsequent non-hierarchical k-means clustering to derive CTs for the analysis of time series of a modified PDSI in Nebraska (USA).

The above-mentioned studies substantiate distinct linkages between drought and large-scale atmospheric circulation in general and CTs in particular, pointing out the suitability of CTCs to capture circulation characteristics relevant for the spatiotemporal variability of droughts. However, so far, there have been no systematic attempts to determine CTCs that, among the multitude of classification methods, are particularly suited to capture drought variability in different regions.

Against the background of the current state of research outlined above, the main objectives of the present study are:

- to analyse the quantitative relationships between CTs and interannual SPI variations in Central Europe in the period 1950 to 2010,
- to establish statistical models for the estimation of gridded SPI values from occurrence frequencies of CTs derived from varying CTCs and
- to identify the most suitable approaches for the statistical downscaling of possible future SPI variations

Recent analyses have shown that the relevance of circulation type classifications for surface climate strongly varies depending on the general classification approach and—partly even more distinctly—the specific configuration settings, concerning for example the number of types or the size of the spatial domain (e.g. Beck and Philipp 2010, Beck et al. 2013, Beck et al. 2014).

Consequently different variants of CTCs comprising various approaches and configuration settings are employed in statistical models that are developed for linking frequency changes of CTs to time series of gridded SPI by multiple linear regression analyses. The skill of the resulting models is evaluated in order to allow for their quantitative comparison. In this way, the most suitable approaches for investigating circulation-drought relationships—which in future studies

may be utilized for the statistical downscaling of possible future SPI variations - are determined. Although summer droughts are in general most relevant in Central Europe, droughts may appear and may have strong impact in the other seasons as well. Therefore, analyses are run for representative months from all four 3-month seasons. Focusing on the meteorological aspects of drought, from the available multiple SPI time scales the 3-month variant is used in this study.

In Section 2, the data sets and the methodological approaches for circulation type classification and for relating CTs to SPI by multiple linear regression analyses are introduced. Section 3 illustrates the main results. Finally, discussion and conclusions follow in Section 4.

2 Data and methods

2.1 Gridded drought index data

The SPI considers only monthly precipitation data and is based on the standardized probabilities for certain precipitation amounts that are estimated from long-term observed precipitation data. Due to the fact that monthly precipitation most often is not normally distributed the calculation of the SPI involves the fitting of an adequate theoretical distribution to the precipitation data and the subsequent equiprobability transformation to the standard normal distribution (McKee et al. 1993). Resulting SPI values are negative for dry and positive for wet conditions. Table 1 gives the SPI value intervals and the event probabilities for different SPI categories.

Gridded SPI data are retrieved from the data set provided by the CliSAP-Integrated Climate Data Center (ICDC) at the University of Hamburg (Sienz et al. 2012). This SPI data set is based on the European Climate and Data Assessment (ECA&D), E-OBS gridded data set Version 4.0 (Haylock et al. 2008) with a 0.5° by 0.5° horizontal resolution. It covers the region 11 W to 60 E and 35 N to 72 N, with 0.5° resolution in both directions and is available for the period January 1951 to December 2010 in monthly resolution for the time scales 1, 3, 6, 9, 12, 24 and 48 months. Sienz et al. (2012) modify the SPI calculation according to McKee et al. (1993) by fitting the Weibull distribution instead of the gamma distribution to the precipitation data. As Sienz et al. (2012) show, this leads to more reliable estimates of extremely dry and wet conditions (see Sienz et al. (2012) for further details on the calculation and the characteristics of the improved SPI). Although the most common utilization of the SPI is to describe extreme drought events, it can in general be used to quantify dry and as well wet conditions of varying strength.

Table 1 Definition of the standardized precipitation index (SPI). SPI intervals, category labels and event probabilities

SPI value	SPI category	Probability [%]
SPI≥2	Extremely wet	2.3
2>SPI≥1.5	Severely wet	4.4
1.5>SPI≥1	Moderately wet	9.2
1>SPI>−1	Normal	68.2
−1≥SPI>−1.5	Moderately dry	9.2
−1.5≥SPI>−2	Severely dry	4.4
SPI≤−2	Extremely dry	2.3

Focusing on Central Europe and on meteorological aspects of dryness/wetness, SPI data calculated for the time scale of 3 months within the region 5.5 E to 16 E and 46.5 N to 55 N are selected for the present analyses. To consider all seasons of the year, 3-month SPI data for February, May, August and November are used.

2.2 Circulation data and circulation type classifications

For the determination of CTs via the different classification approaches described below, 2.5° by 2.5° gridded daily 12 UTC sea level pressure (SLP) data from the NCEP/NCAR reanalysis 1 data archive (Kalnay et al. 1996) for the period covered by SPI data, 1951 to 2010, have been used.

Daily CTs have been derived applying several variants of two different classification methods to the SLP data using the cost733class classification software (Philipp et al. 2010, 2014). In more detail, the two selected CTCs are:

The so-called Grosswettertypes (GWT) or Prototype classification (Beck et al. 2007) that arranges cases (daily SLP fields) into classes (circulation types) according to the main direction of the large-scale flow and the cyclonicity. To this end, spatial correlation coefficients between daily SLP fields and three prototypical flow patterns representing idealized zonal (westerly), meridional (northerly) and cyclonic flow conditions respectively are calculated (denoted as coefficients of zonality (Z), meridionality (M), and vorticity (V) hereinafter). Circulation types are then defined as particular ideal combinations of these coefficients (see Table 2). Finally, each daily SLP field is assigned to one circulation type according to the following criteria: (1) a maximum absolute value of V leads to a pure cyclonic (positive sign) or anticyclonic (negative sign) type respectively; (2) remaining cases (daily SLP fields) are assigned to the most similar (in terms of the minimum Euclidean distance of the respective Z and M coefficients) directional circulation type; directional types may further be subdivided according to (3) the sign of the V coefficient into cyclonic (positive sign) and anticyclonic (negative sign) subtypes or according to (4) values of the standardized V coefficient into cyclonic (standardized V>0.42), indifferent (−0.42≤standardized V≤0.42) and anticyclonic (standardized V<− 0.42) subtypes. Depending on how many criteria are considered, varying numbers of classes (circulation types)

Table 2 Overview of classification criteria and resulting circulation types for variants of the Grosswettertypes (GWT) classification (Beck et al. 2007) comprising 10, 18 and 27 circulation types respectively. Pure cyclonic/anticyclonic circulation types are denoted by letters L (low pressure), H (high pressure) and I (indifferent) respectively. Acronyms for directional circulation types include information on flow direction (W=west, N=north, E=east, S=south; and derived combinations) and on cyclonicity (c=cyclonic, i=indifferent, a=anticyclonic; note that the definition of c and a differs between the 18 and 27 type variants of the classification scheme). *Z*, *M*, *V* denote the zonality, meridionality and vorticity coefficients respectively (see text for further explanation); SV is the standardized vorticity coefficient and d is the euclidean distance between an object (daily SLP field f) to classify and a certain main directional type in the two-dimensional feature space spanned by the respective *Z* and *M* coefficients (e.g. d(W)=SQRT((Z_f-ZW)**2+(M_f-MW)**2))

		10 Types Central low/high pressure types		18 Types		27 Types
Criteria (1)						
$V_f=\max(Z_f, M_f, V_f)$	$V_f>0$	L		L	$SV_f>0.42$	L
					$-0.42\leq SV_f\leq 0.42$	I
$V_f=\max(Z_f, M_f, V_f)$	$V_f<0$	H		H	$SV_f<-0.42$	H
		Main directional types (ideal combinations of *Z* and *M* coefficients)		Cyclonic/anticyclonic subtypes		Cyclonic/indifferent/ anticyclonic subtypes
Criteria (1)	Criteria (2)		Criteria (3)		Criteria (4)	
			$V_f>0.0$	Wc	$SV_f>0.42$	Wc
$V_f\neq\max(Z_f, M_f, V_f)$	d(W)=min(d(W),…, d(SW))	W (Z_W=1.0, M_W=0.0)			$-0.42\leq SV_f\leq 0.42$	Wi
			$V_f<0.0$	Wa	$SV_f<-0.42$	Wa
			$V_f>0.0$	NWc	$SV_f>0.42$	NWc
$V_f\neq\max(Z_f, M_f, V_f)$	d(NW)=min(d(W),…, d(SW))	NW (Z_{NW}=0.7, M_{NW}=0.7)			$-0.42\leq SV_f\leq 0.42$	NWi
			$V_f<0.0$	NWa	$SV_f<-0.42$	NWa
			$V_f>0.0$	Nc	$SV_f>0.42$	Nc
$V_f\neq\max(Z_f, M_f, V_f)$	d(N)=min(d(W),…, d(SW))	N (Z_N=1.0, M_N=0.0)			$-0.42\leq SV_f\leq 0.42$	Ni
			$V_f<0.0$	Na	$SV_f<-0.42$	Na
			$V_f>0.0$	NEc	$SV_f>0.42$	NEc
$V_f\neq\max(Z_f, M_f, V_f)$	d(NE)=min(d(W),…, d(SW))	NE (Z_{NE}=−0.7, M_{NE}=0.7)			$-0.42\leq SV_f\leq 0.42$	NEi
			$V_f<0.0$	NEa	$SV_f<-0.42$	NEa
			$V_f>0.0$	Ec	$SV_f>0.42$	Ec
$V_f\neq\max(Z_f, M_f, V_f)$	d(E)=min(d(W),…, d(SW))	E (Z_E=−1.0, M_E=0.0)			$-0.42\leq SV_f\leq 0.42$	Ei
			$V_f<0.0$	Ea	$SV_f<-0.42$	Ea
			$V_f>0.0$	SEc	$SV_f>0.42$	SEc
$V_f\neq\max(Z_f, M_f, V_f)$	d(SE)=min(d(W),…, d(SW))	SE (Z_{SE}=−0.7, M_{SE}=−0.7)			$-0.42\leq SV_f\leq 0.42$	SEi
			$V_f<0.0$	SEa	$SV_f<-0.42$	SEa
			$V_f>0.0$	Sc	$SV_f>0.42$	Sc
$V_f\neq\max(Z_f, M_f, V_f)$	d(S)=min(d(W),…, d(SW))	S (Z_S=0.0, M_S=−1.0)			$-0.42\leq SV_f\leq 0.42$	Si
			$V_f<0.0$	Sa	$SV_f<-0.42$	Sa
			$V_f>0.0$	SWc	$SV_f>0.42$	SWc
$V_f\neq\max(Z, M, V)$	d(SW)=min(d(W),…, d(SW))	SW (Z_{SW}=0.7, M_{SW}=−0.7)			$-0.42\leq SV_f\leq 0.42$	SWi
			$V_f<0.0$	SWa	$SV_f<-0.42$	SWa

result from the classification scheme. For the analyses presented here, three variants comprising 10, 18 and 27 circulation types, respectively, have been developed. An overview of the classification scheme and the three resulting variants of the GWT classification used in this contribution is given in Table 2. Exemplary SLP composites for circulation types

resulting from the basic classification into 10 types are displayed in Fig. 1.

The second classification approach is based on the non-hierarchical k-means clustering algorithm (e.g. Hartigan 1975). For the variant used here, the most dissimilar cases (daily SLP fields) included in the data set are used to determine the initial starting partition, as suggested by Enke and Spekat (1997) and provided in a corrected form by Philipp et al. (2010). In iterative steps, it is then evaluated for each object (daily SLP field) whether it already is in the most similar cluster (in terms of the Euclidean distance between the object and the respective cluster centroid) or whether it has to be assigned to another—nearer—cluster (which in turn leads to a modification of the respective cluster centroids). This iterative process is stopped when all objects are assigned to their nearest cluster and further re-assignments lead to no further improvement (in terms of reduction of the within-cluster variances). In accordance to the GWT classification, this k-means classification based on highly differentiated starting partitions (DKM) has also been run for 10, 18 and 27 classes (circulation types).

The GWT and DKM classifications used in this investigations have been selected from the large pool of available CTCs (e.g. Huth et al. 2008, Philipp et al. 2010, Philipp et al. 2014),

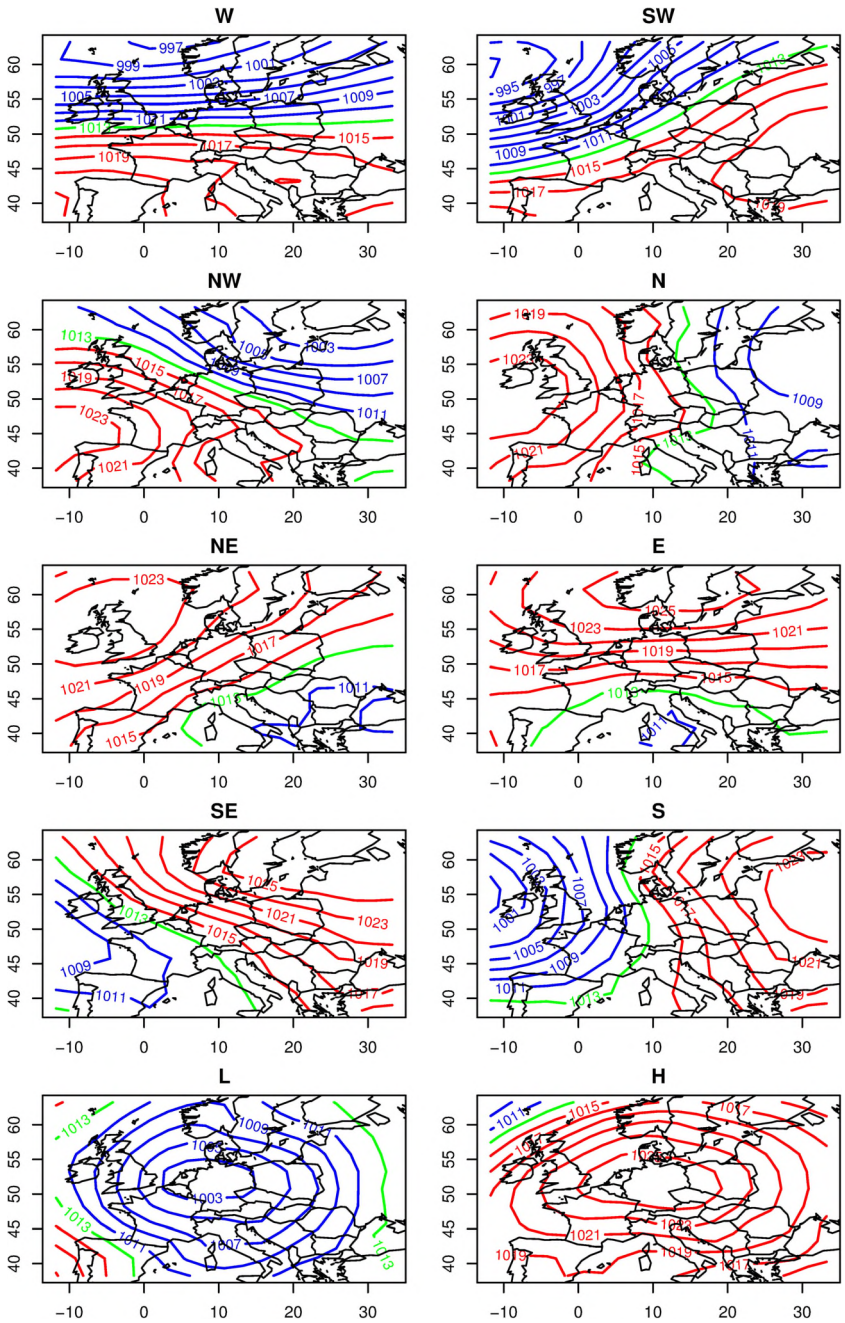

Fig. 1 SLP (hPa) composites for 10 circulation types resulting from the GWT classification applied to daily SLP (1950–2010) for the domain 13° W–34.5° E/37°–64.5°N

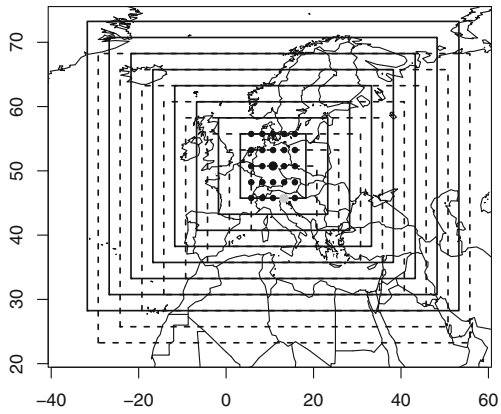

Fig. 2 Sizes and locations of spatial domains used for deriving circulation type classifications (CTCs). *Dots* indicate 25 central points of spatial domains. As examples solid and dashed rectangles depict eight varying domain sizes centered on 5.75° E/50.75° N (*bold black dot*) and 8.25° E/45.75° N (*bold grey dot*) respectively

because previous studies (e.g. Beck and Philipp 2010, Huth 2010, Schiemann and Frei 2010) have shown that in many cases threshold based classifications (e.g. GWT classification) and classifications based on optimisation algorithms (e.g. DKM classification) are most appropriate to resolve precipitation variations in central parts of Europe.

It has furthermore been shown (e.g. Beck and Philipp 2010, Huth 2010) that the number of classes (CTs) has distinct effects on the discriminative power (synoptic skill) of circulation type classifications for surface climate data, with—in most cases—synoptic skill increasing with the number of CTs. Whereas the DKM classification may provide any number of CTs, the GWT classification is applicable only for a limited number of types. Therefore—in order to keep results comparable—both CTCs have been run for 10, 18 and 27 CTs.

The effect of domain size on the synoptic skill of CTCs has recently been reported by Beck et al. (2013). Besides the general finding that the use of smaller spatial domains appears to be more appropriate to capture precipitation variations, Beck et al. (2013) also detected distinct spatial variability in optimal domain size. Not yet analysed in detail is the relevance of the exact location of the domain, either centered over the target location or not. Taking into account these points, CTCs have been applied to spatial domains of varying size and location for the present study. In detail, eight domain sizes have been defined. The smallest domain ranges over 15° in west–east and over 10° in north–south direction, the largest domain spans 85° in west–east and 45° in north–south direction. Such sequences of domains have been defined for 25 locations over Central Europe. An overview of sizes and locations of the spatial domains is given in Fig. 2.

Summing up over all CTC variants, finally 1,200 CTCs have been generated. All classification variants have been produced using the cost733class software package (Philipp et al. 2010, Philipp et al. 2014) that has been developed within the framework of the COST Action 733 "Harmonisation and Applications of Weather Type Classifications for European Regions" (see e.g. Huth et al. 2010 for an overview).

2.3 Multiple linear regression models

For each of the 318 grid cells over land within the domain 5.5 E to 16 E and 46.5 N to 55 N, 3-month SPI values are modeled via multiple linear regression analysis (MRA), using different combinations of subsets of the following set of potential predictors.

Firstly, occurrence frequencies (number of days assigned to a specific CT) are calculated for each CT resulting from the varying CTCs for 3-month periods corresponding to the 3-month SPI data. Time series of CT frequencies are thus available for the two classification methods GWT and DKM, for three different numbers of types, for four 3-month episodes, and for eight domain sizes at 25 locations. Secondly, 3-monthly mean circulation indices of zonality, meridionality and cyclonicity are calculated from respective daily values that are determined by the GWT classification (see Section 2.2). These variables may be considered as proxies for within-type characteristics of CTs, concerning the above-mentioned circulation characteristics—however, averaged over all CTs.

Table 3 Minimum, mean and maximum values of different skill scores (see text for explanation) determined for the 10 % best performing (according to MSSS) multiple regression models for the estimation of gridded 3-month SPI from large-scale circulation

	February			May			August			November		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
R^2	0.14	0.53	0.76	0.12	0.46	0.78	0.08	0.39	0.72	0.16	0.43	0.69
MSSS	0.13	0.53	0.76	0.10	0.46	0.78	0.08	0.38	0.72	0.15	0.43	0.68
HSSdry	−0.06	0.41	0.76	−0.11	0.37	0.77	−0.13	0.32	0.73	−0.13	0.33	0.73
HSSwet	−0.06	0.41	0.72	−0.08	0.38	0.80	−0.17	0.31	0.75	−0.16	0.30	0.80

Table 4 Percentage occurrences of different configuration settings of circulation type classifications (classification method—CM, number of types—NT, size - DS and location—DL of the domain) and multiple regression models (potential predictors—PP, number of predictors—NP) determined for the 10 % best performing (according to MSSS) multiple regression models for the estimation of gridded 3-month SPI from large-scale circulation. GWT and DKM refer to the classification methods described in Section 2.2. GD and LD denote domains centered over 10.75° E/50.75° N and centered as close as possible to the target location respectively (see Section 2.2). CT and CTCI indicate potential predictor sets comprising only circulation type frequencies (CT) or circulation type frequencies and additional monthly circulation indices (CTCI). Numbers in italics indicate percentages that together add up to more than 50% of the respective configuration category

	CM		NT			DS								DL		PP		NP						
	GWT	DKM	10	18	27	1	2	3	4	5	6	7	8	GD	LD	CT	CTCI	1	2	3	4	5	6	7
February	*68*	32	3	32	*65*	11	*16*	*17*	*19*	14	14	7	2	48	*52*	44	*56*	0	1	22	*40*	*27*	9	1
May	42	*58*	13	24	*63*	*18*	*23*	*20*	16	8	6	6	3	*51*	49	32	*68*	0	6	22	*34*	*25*	11	2
August	*54*	46	20	*33*	*47*	11	*40*	*20*	11	6	6	4	2	*58*	42	17	*83*	1	22	*42*	*24*	8	3	1
November	*68*	32	10	28	*62*	*37*	*25*	19	10	3	2	2	2	*53*	47	26	*74*	0	7	*22*	*46*	21	3	1

For each model, CT frequencies from one CTC variant are utilized as potential predictors, either exclusively or in combination with the three circulation indices. Thereby, both CT frequencies and circulation

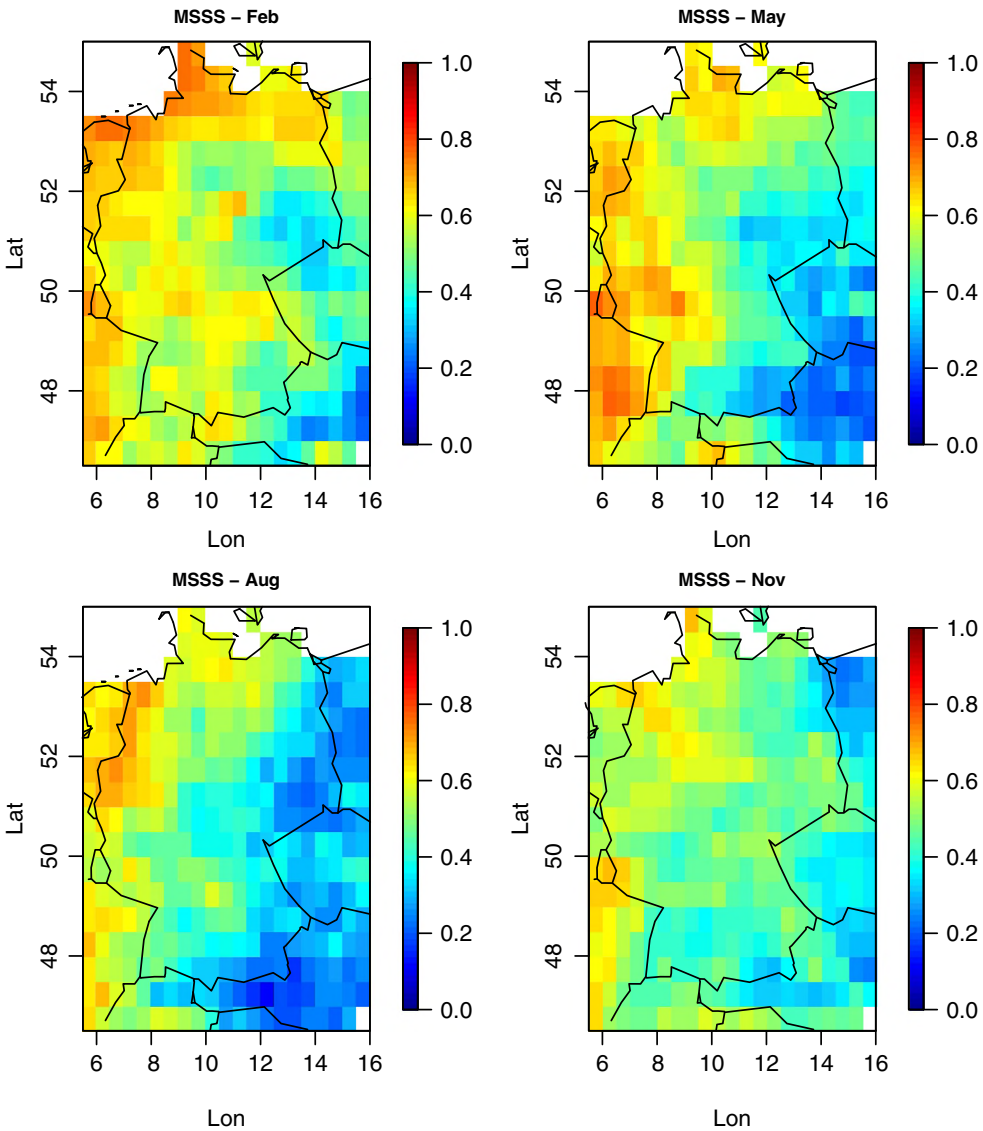

Fig. 3 Spatial distribution of the Mean Square Skill Score (MSSS) of the best performing multiple regression model for each grid cell in February, May, August and November

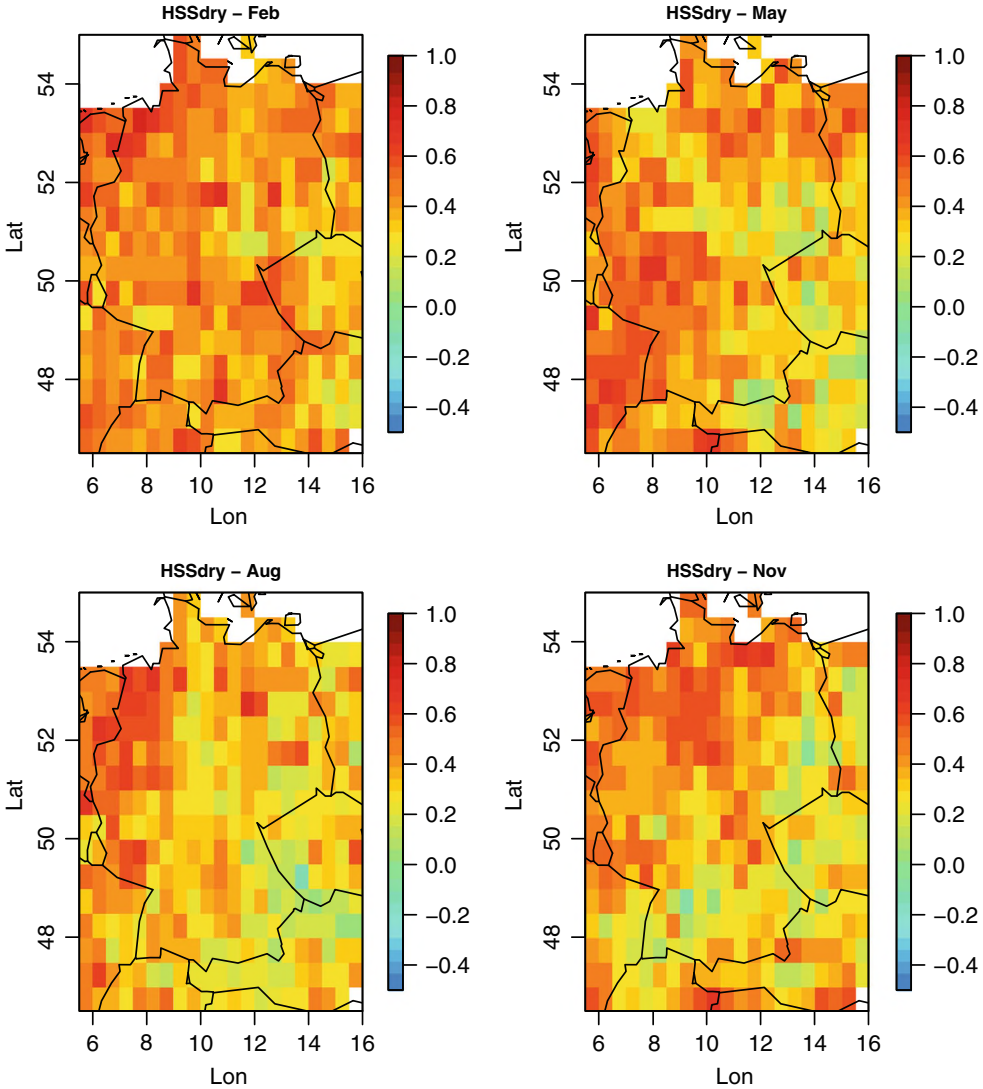

Fig. 4 Spatial distribution of the Heidke Skill Score estimated for the combined SPI event categories severely and extremely dry (HSSdry) of the best performing (in terms of MSSS) multiple regression model for each grid cell in February, May, August and November

indices, are either derived for a domain centered on 10.75 E/50.75 N or for a domain centered on the grid point which is closest to the location of the predictand SPI series. Hence, for each predictand variable (SPI series at a specific grid location), 192 different potential predictor combinations can be defined (2 classification methods × 3 numbers of types × 2 domain locations × 8 domain sizes × 2 variants including or excluding monthly mean circulation indices). To avoid the influence of trends in the data on the model skill, the linear trend has been removed from all predictor and predictand time series.

Analysing the relationship between CTs/circulation indices and SPI via MRA involves several steps which are further detailed below.

In order to find adequate sets of predictor variables, firstly stepwise MRA is applied to 100 random samples of 40 years each (thus leaving 20 years unused). These random samples have been defined once and are used for the predictor screening for all models. Model selection in stepwise MRA in each case is based on the Akaike information criterion (AIC) (Akaike 1974).

Based on the results of the 100 explorative stepwise MRA runs, the most frequent number of predictor variables and as well the most often chosen predictor variables are determined. Thereby, only those models are considered for which the basic assumptions underlying the MRA are fulfilled. These assumptions are in particular normality, homoscedasticity and independence of residuals which have been tested using the Shapiro-Wilk test (e.g. Royston 1982), the Breusch-Pagan test (e.g. Krämer and Sonnberger 1986) and the Durbin-Watson test (e.g. Krämer and Sonnberger 1986), respectively. For rejection of the null hypothesis, the significance level alpha has been set to 0.05 for all three tests.

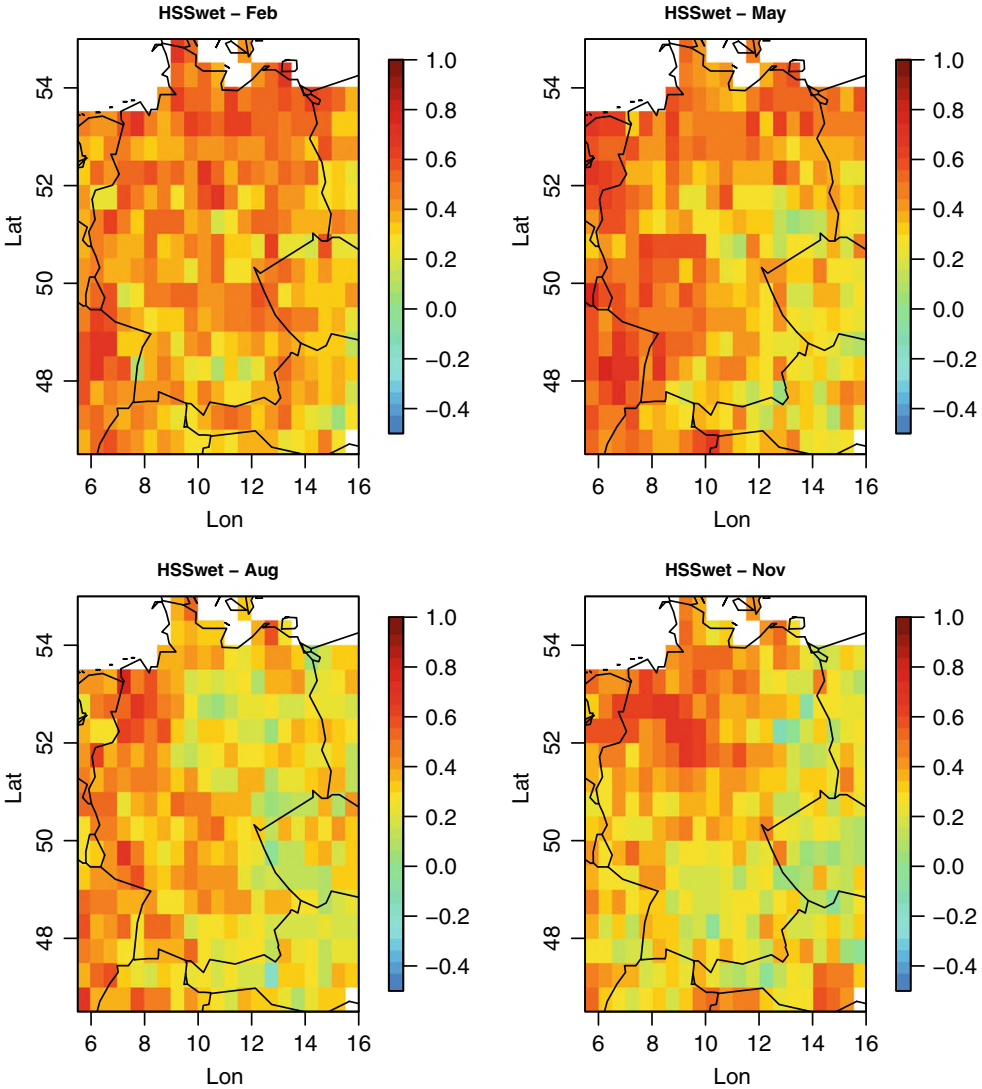

Fig. 5 Spatial distribution of the Heidke Skill Score estimated for the combined SPI event categories severely and extremely wet (HSSwet) of the best performing (in terms of MSSS) multiple regression model for each grid cell in February, May, August and November

Furthermore, in case of dependence between the predictors—i.e. significant (for alpha=0.01) correlations between predictors—only the one predictor variable showing the highest correlation coefficient to the predictand is kept, while the others are removed from the set of potential predictors.

As a result of the screening procedure, a reduced set of predictors is defined for each of the 192 MRA settings at each of the 318 SPI grid locations. These fixed combinations of predictors are used in subsequent analyses without further rearrangements.

Utilizing the previously specified predictor variables, MRA is applied to predictand and predictor data for the whole period from 1951 to 2010. Based on these MRAs, the importance of each CT or circulation index for SPI variability is estimated for each MRA. For this purpose, the relative importance metric LMG (referring to author names Lindeman, Merenda and Gold) as proposed by Lindeman et al. (1980) has been calculated according to Grömping (2007). LMG decomposes the explained variance of each model and provides estimates of the percentage contributions from the different predictor variables.

Finally, a leave-one-out cross-validation procedure has been performed to achieve reliable evidence of the robustness of the different models. To this end, linear multiple regression models (MRM) using the screened predictor variables have been constructed for calibration periods, each comprising all but 1 year of the period 1951–2010. Thus, advancing from 1951 to 2010 SPI values for each year are modeled by applying the MRM fitted to the remainder of all years to predictor data for the omitted year.

For quantifying the skill of the different MRMs, several metrics have been estimated on the basis of the observed and the modeled SPI series.

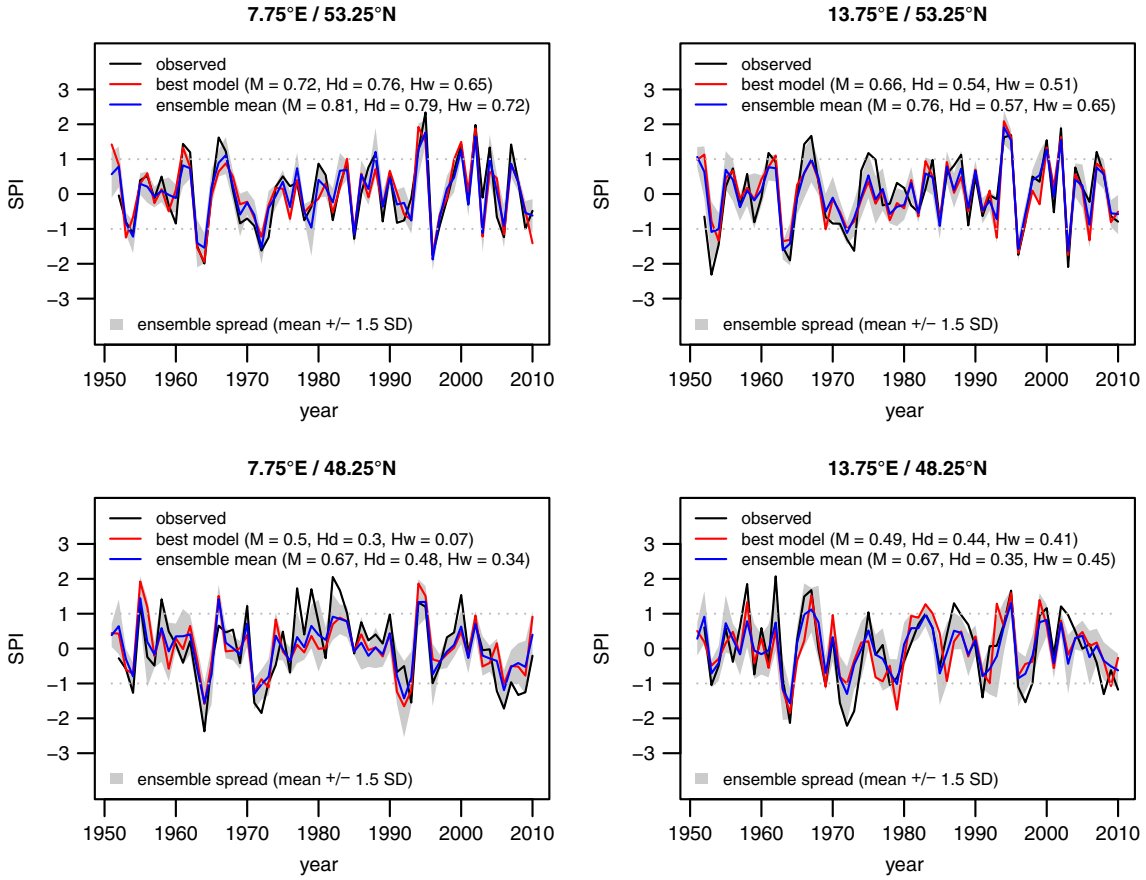

Fig. 6 Time series of observed and modeled 3-month SPI time series for February, for four selected grid cells (respective latitude and longitude coordinates are indicated in the title of each plot). In addition to the modeled series resulting from the best performing model (best model), the mean of the 10 % best performing models (ensemble mean) and the interval ±1.5 times the respective standard deviation around the mean are given (ensemble spread). Horizontal dotted grey lines indicate the threshold for the moderately dry/wet SPI category respectively. The skill scores MSSS (M), HSSdry (Hd) and HSSwet (Hw) estimated for the best model and the ensemble mean are indicated in brackets in the top left legend of each time series plot

From the range of accuracy measures that may be used for the comparison of observed and modeled series of continuous variables (e.g. Wilks 2006), the squared Pearson correlation coefficient (R^2) between observed and modeled SPI series and the mean square skill score (MSSS, Murphy 1988) have been calculated.

MSSS is based on the mean squared error (MSE) estimated for the model and for the climatological reference, respectively:

$$\mathrm{MSSS} = 1 - \frac{\mathrm{MSE}}{\mathrm{MSE}_{\mathrm{Clim}}}$$

With

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - a_i)^2$$

and

$$\mathrm{MSE}_{\mathrm{Clim}} = \frac{1}{N}\sum_{i=1}^{N} \left(\overline{a} - a_i\right)^2$$

Thereby N is the number of cases, y_i is the modeled value for case i, a_i is the respective observed value and a is the mean over all a_i. MSSS values less than or equal 0 imply a model quality worse or equal to climatology, while a maximum MSSS value of 1 indicates a perfect model.

To take into account the particular relevance of pronounced dry and wet events, the ability of the models to reproduce such events has been checked by applying the Heidke skill score (HSS). HSS has been calculated according to Wilks (2006) as

$$\mathrm{HSS} = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}$$

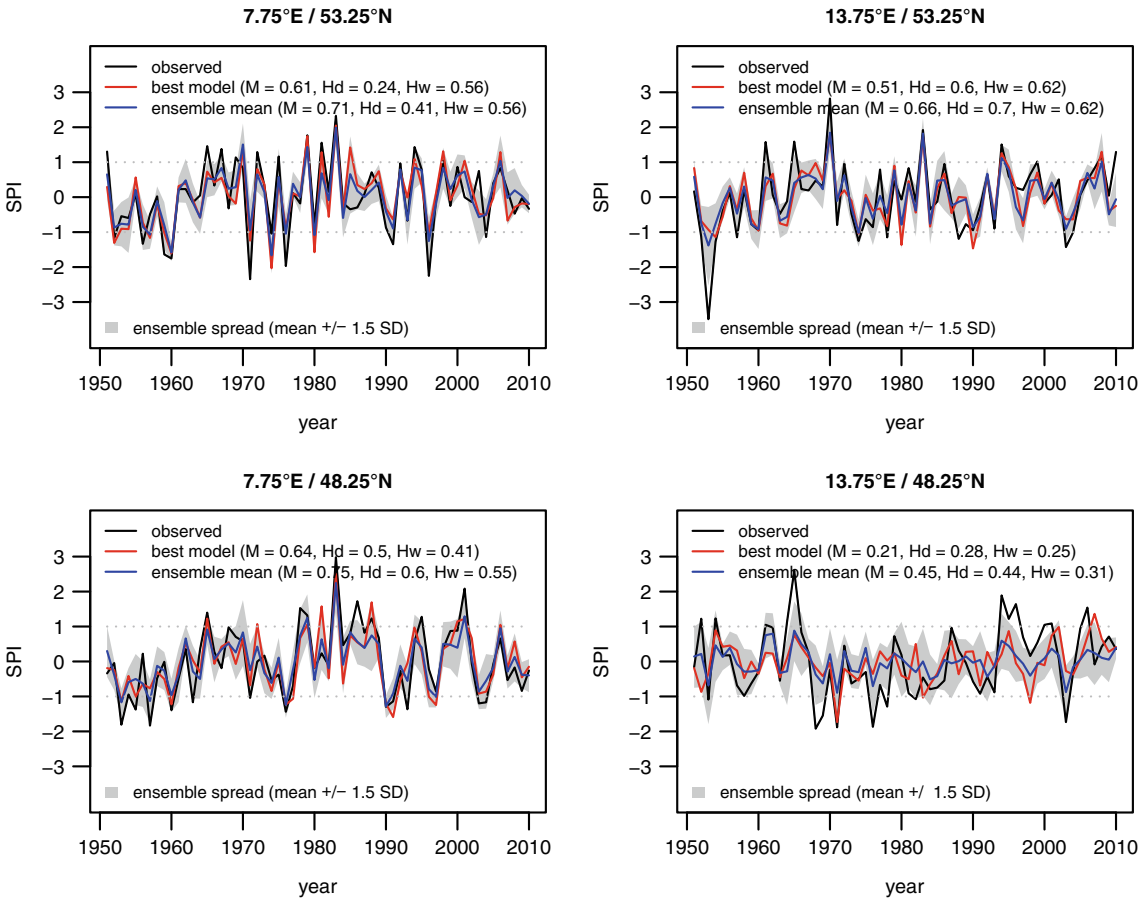

Fig. 7 Same as Fig. 6 but for May

Thereby a is the number of observed events that are correctly modeled (hits), b is the number of modeled events that are not observed (false alarms), c is the number of events that are observed but not modeled (misses), and d is the number of non-events in both observation and model (correct negative). HSS values<0 indicate no skill while a value of 1 implies a perfect model. HSS has been estimated separately for two SPI event categories (see Table 1), defined by an SPI<−1.0 (moderately to extremely dry, HSSdry) or an SPI>1.0 (moderately to extremely wet, HSSwet), respectively.

3 Results

According to the main purpose of this study—the determination of CTC-based regression models that are best suited to capture the relevant links between the atmospheric circulation and drought index variations—the presentation of results focuses on a selection of those models that perform best in terms of model skill. For each grid cell, the 20 best performing multiple regression models have been selected from the 192 available models according to the MSSS, estimated on the basis of the leave-one-out cross-validation. This selection of the roughly 10 % best models for each grid cell leads to model "ensembles" which establish the basis for depicting the main characteristics of model configurations that are most suitable for estimating 3-month SPI from the large-scale atmospheric circulation.

3.1 Overall skill of best performing models

Minimum, mean and maximum values, respectively, of the skill scores (see Section 2.3) estimated over all SPI grid cells are given in Table 3.

A first remarkable finding from Table 3 is that the values for R^2 and MSSS are almost equal. This points to the fact that both, the unconditional—or overall or systematic—bias (difference between the model mean and the observation mean) and the conditional bias (correlation coefficient minus the ratio of the variances of model and observation) of the models are very small (see Murphy 1988 concerning the decomposition of the MSSS). Thus, instead of R^2 and MSSS, mainly MSSS is further discussed in the following.

Highest and least mean skill in terms of MSSS is reached in February and August, respectively while May and November

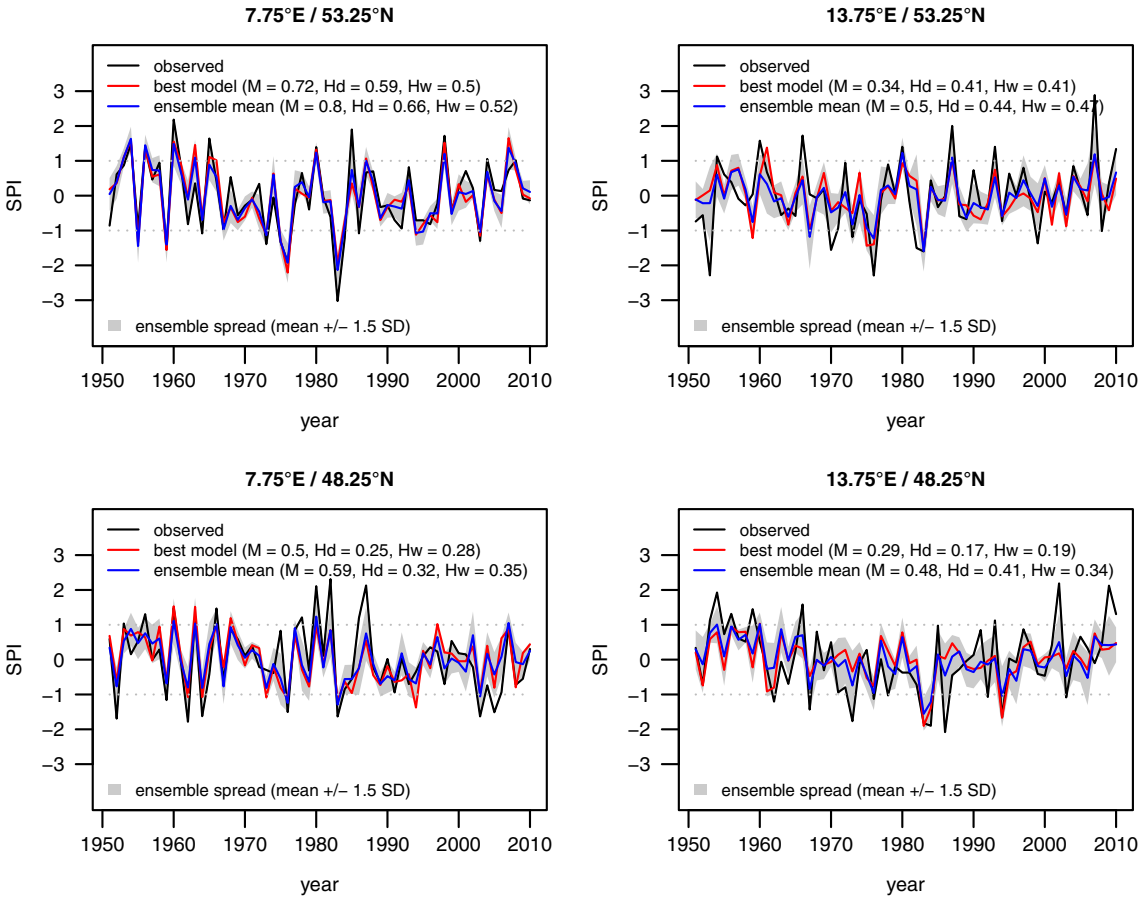

Fig. 8 Same as Fig. 6 but for August

exhibit intermediate mean values of MSSS in general. This points to the fact that the influence of the large-scale circulation on precipitation is most pronounced in winter. On the other hand, interseasonal differences decrease for the maximum and minimum MSSS values out of all single grid cells, indicating that in each month there are at least some grid cells for which very low or rather high skill is reached. Maximum MSSS (and R^2) in all months reaches values between 0.68 and 0.78, implying amounts of SPI variance explained by the respective models of up to 78 %. Minimum values of MSSS—although being rather low—remain positive in all cases, thus indicating at least slightly better skill than climatology for all grid cells.

HSSdry and HSSwet focusing on the ability of the models to capture negative and positive SPI deviations, respectively, indicate higher mean skill in February and May compared to August and November. Maximum values of both HSS variants exceed 0.7 in all months. Together with negative minimum values appearing in all months and for both HSS variants, this indicates a distinct variation in the ability of the models to reproduce dry and wet anomalies, respectively.

3.2 Preferred configuration settings of best performing models

Besides the overall skill of the best performing models, it is of interest in how far these models show preferred configuration settings concerning the underlying CTCs and the multiple regression analyses. Table 4 depicts the appearance (given as percentages) of various configuration settings in the selected best performing MRMs.

With respect to the general classification method, it becomes apparent that—with the exception of May—the GWT classification appears more frequent than the DKM classification. Most probably this is due to the fact that the GWT classification variants comprising 18 and 27 types, respectively, explicitly differentiate between cyclonic and anticyclonic CTs.

The spatial distribution of GWT/DKM preference (not shown) exhibits no interpretable structures.

In the majority of cases, CTCs comprising 18 or 27 CTs are included in the best performing models. This predominance is least obvious in August. It is also the only month for which some—however, only hardly detectable—spatial patterns of

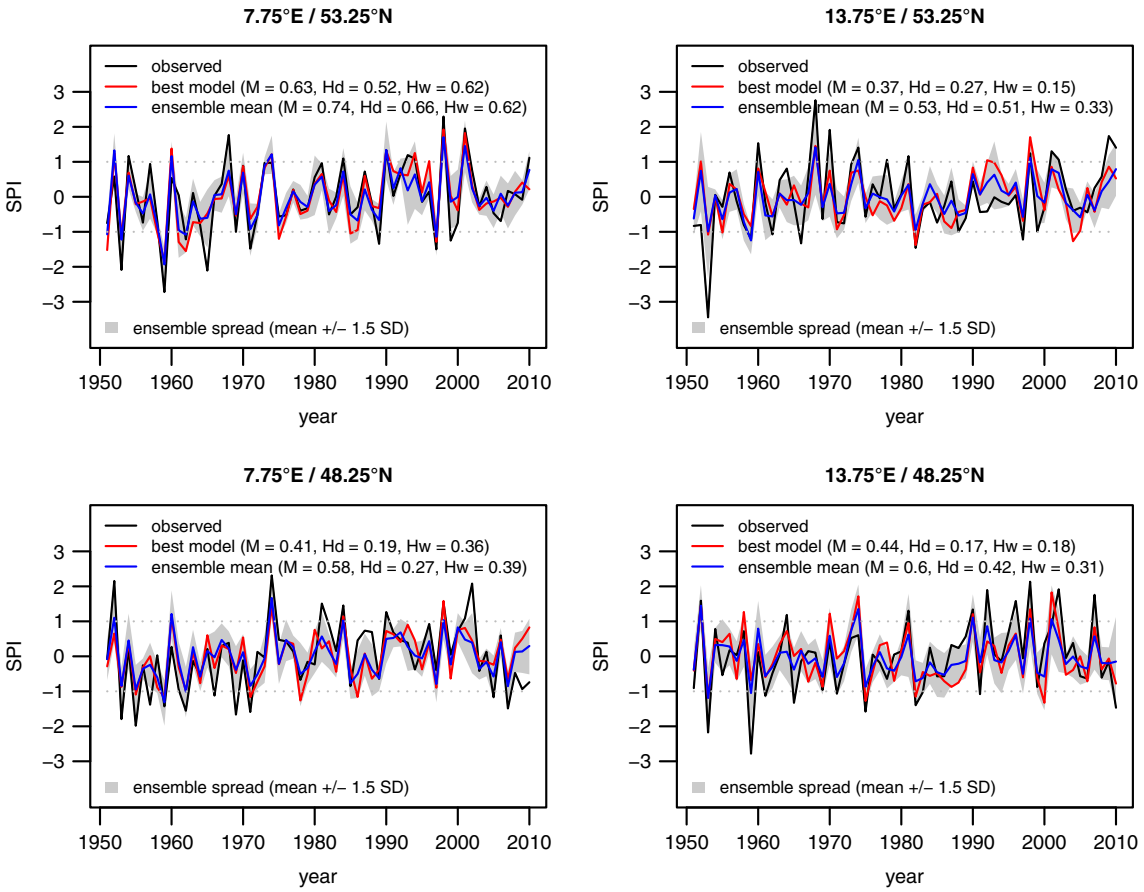

Fig. 9 Same as Fig. 10 but for November

preferred CT numbers, with less types in the west/northwest of the domain, can be seen (not shown).

CTCs applied to smaller domains are more frequently chosen in the best performing models than CTCs applied to larger spatial domains. This preponderance is less distinct in February which is in accordance with findings from Beck et al. (2013). However, again, no clearly structured spatial patterns concerning the preferred domain size become obvious (not shown).

With respect to the location of the domain relative to the target grid cell, no clear-cut preferences can be deduced from Table 4. This is also true for the respective spatial distribution (not shown). Obviously, shifts in the location of the domain—at least within the range studied here—are not as relevant as variations in the size of the domain.

A far more distinct difference in percentage occurrences becomes obvious with respect to the preferred sets of potential predictors utilized in the best performing models. Clearly, the majority of the best performing models relies on sets of predictors that include mean indices for zonality, meridionality and vorticity in addition to occurrence frequencies of circulation types. This appears most pronounced in August and least pronounced in February. These seasonal differences may be explained by the larger amount of climatic within-type variability in summer that in turn is partly due to corresponding variations in circulation characteristics, reflected in the presently used circulation indices (Beck et al. 2007). Again, no interpretable respective spatial patterns occur.

Finally, the selected best performing models may be partitioned with respect to the number of predictors that are chosen for the final model. In this regard, it can be stated that the majority of models comprises between three and five predictors. Fewest predictors are used in August. The respective spatial distributions (not shown) exhibit a tendency towards a decrease in the number of predictors from the northwest to the southeast of the Central European domain.

3.3 Spatial variations in model skill

In addition to variations in model skill between months, distinct differences in skill are obvious between different locations. The spatial distribution of skill in terms of MSSS of the best performing model approach for each grid cell is depicted in Figs. 3, 4 and 5 illustrate spatial variations of HSSdry and HSSwet, respectively. Time series of observed

Table 5 Frequencies of appearance of circulation types and mean circulation indices as leading predictors (Pred 1 to Pred 4) in the best GWT based multiple regression models for 3-month SPI at each grid cell and respective mean relative importance (explained variation in %) estimates. Only those 14 circulation types and mean circulation indices appearing most often as a leading predictor in the best models are shown. Numbers in italics indicate the two highest values in each row

		Cyclonic types					Anticyclonic types				Indifferent types			Indices	
		Wc	SWc	NWc	NEc	L	Na	NEa	Ea	H	Wi	NWi	Si	*V*	*Z*
Frequency of appearance															
Feb	Pred 1	*197*	6	16		1		2		5	*64*	11	6	2	1
	Pred 2-4	82	74	*97*	14	17	14	57	5	*83*	40	59	47	54	46
May	Pred 1	51	1	5	16	18	23	4	*55*	25	3	1		*60*	21
	Pred 2-4	*110*	41	41	38	65	51	44	*66*	46	24	13	18	57	38
Aug	Pred 1	12	6	7	9	8	8	18	11	*18*	2			*194*	
	Pred 2-4	*79*	28	62	42	50	29	34	58	*66*	21	13	4	21	14
Nov	Pred 1	35	2	*48*	12	5		1	1	6	36		4	*141*	1
	Pred 2-4	*139*	23	97	26	58	37	16	17	39	*116*	13	29	68	11
Mean relative importance															
Feb	Pred 1	*33.0*	21.5	25.6		20.3		15.1		15.7	28.0	21.3	17.0	*32.0*	18.0
	Pred 2-4	*12.7*	9.2	10.5	8.3	9.7	3.5	9.5	4.2	8.8	*11.3*	9.7	9.4	10.2	11.1
May	Pred 1	22.8	17.5	16.0	15.6	*24.4*	17.8	16.1	21.2	17.4	*26.1*	14.2		21.5	22.3
	Pred 2-4	11.3	11.2	10.8	7.8	11.1	10.1	*12.0*	10.8	9.1	10.1	9.8	9.0	11.7	*12.1*
Aug	Pred 1	18.6	20.0	*23.3*	16.7	20.7	16.3	17.6	18.2	19.9	13.9			*31.1*	
	Pred 2-4	9.6	8.7	9.8	8.1	8.8	*10.0*	8.5	9.7	*10.4*	7.9	*10.0*	7.4	6.9	8.1
Nov	Pred 1	20.5	*22.1*	21.6	19.9	17.9		12.1	12.6	19.9	20.7		15.7	*23.0*	14.7
	Pred 2-4	11.2	9.5	11.2	11.4	11.1	8.9	8.9	8.7	10.2	10.9	7.9	9.6	*11.9*	*13.3*

and modeled 3-month SPI (best performing model, mean over the 20 best models and respective range ±1.5 times the standard deviation) for selected grid cells are shown in Figs. 6, 7, 8 and 9.

Starting with spatial variations in MSSS (Fig. 3), a pronounced decrease in skill can be stated from the western/northwestern parts to the eastern/southeastern parts of the Central European domain, explicable by the decreasing exposure to synoptic-scale cyclonic systems towards the more continental regions. The general spatial gradient in skill is visible in all months, but appears most pronounced in February and least distinct in November. Main features of the spatial variations detected for MSSS are as well discernible for HSSdry (Fig. 4) and HSSwet (Fig. 5).

These patterns of spatial variations in model skill also show up from time series plots in Figs. 6,7, 8 and 9. Grid cells for which time series are presented have been selected in order to represent the northwestern, the northeastern, the southeastern and the southwestern regions of the Central European domain featuring distinctly varying model skill.

Notably high skill (in terms of MSSS above 0.6 for the best model) is observed in all months for the northwestern grid cell and as well for the northeastern grid cell in February and for the southwestern grid cell in May. Distinctly low skill with values of MSSS (for the best model) less than or equal 0.3, on the other hand, is apparent for the southeastern grid cell in May and August. Connected with higher/lower skill of the best model for different grid cells, respective less/more variations among the ensemble of the 20 best models (indicated by the grey shaded intervals around the ensemble mean) can be stated.

A further remarkable finding is illustrated in Figs. 6, 7, 8 and 9. In all cases shown in these figures, those SPI series that have been determined by averaging the modeled series resulting from the 20 best models (ensemble mean) reach considerably higher skill (in terms of MSSS) than the modeled SPI series from the best performing model. This increase in skill, when using the ensemble mean instead of the best single model, can (with only a few exceptions) also be stated for the skill scores HSSdry and HSSwet (see Figs. 6, 7, 8 and 9) and in general as well (not shown) for the majority of Central European grid cells used in this study.

With respect to the reproduction of dry and wet anomalies beyond SPI values of −1/+1, Figs. 6, 7, 8 and 9 depict that, while the timing of most of these events is quite well captured by the models, their amplitudes—in particular those of the more severe events—are often underestimated, particularly for those grid cells and months also exhibiting low skill in terms of MSSS.

3.4 Relative importance of predictors

Finally, the most relevant predictors used in the MRMs and respective estimates of their relative importance are depicted in Table 5 and in Figs. 10, 11, 12 and 13.

For the 14 predictors (circulation types and mean circulation indices) appearing most often among the four leading predictors in the best MRM of each grid cell, Table 5 summarizes frequencies of appearance among the four leading predictors and gives estimates of their respective mean relative importance (explained variance). Figures 10, 11, 12 and 13 show estimates of the relative importance of the leading predictor in the best performing (according to MSSS) MRM for each grid cell, as amounts of explained variance (in percent) that can be attributed to the respective predictor.

Two points concerning the presentation of relative importance of leading predictors need to be clarified: First—importance has been estimated on the basis of MRAs applied to the whole period from 1951 to 2010. Thus, the explained variances depicted in Figs. 10, 11, 12 and 13 and in Table 5 must not be related directly to respective R^2 values estimated for the leave-one-out cross-validation. Second—in order to enable a comprehensible and consistent characterisation of the most important predictors, importance estimates for each grid cell are shown for the respective best model that is based on a variant of the GWT classification. This has the advantage that in all cases somewhat familiar denotations of circulation types (see Table 2) appear (instead of arbitrary numbers—as resulting from the KM classification—that may be related to differing circulation types at varying grid cells). The main purpose of this section is simply to highlight which general circulation characteristics, reflected by CTs (and mean circulation indices), are the most important predictors in which months and at which locations. Furthermore, it is intended to show how main parts of the total explained variance are distributed among these leading predictors. Thus, the above-mentioned restrictions should not significantly impair the main findings presented below.

From Table 5, it can be seen that the mean vorticity index (*V*) and several CTs (Wc, NWc, H and Wi) can be denoted as most important predictors, according to their frequencies of appearance among the first four predictors in the best GWT based MRMs. Particularly *V*/Wc appear exceptionally often as the leading predictors in August and November/February. In these months they are the leading predictors for almost or even more than the half of all grid cells, respectively. A distinct differentiation between months emerges according to aggregated appearance frequencies of two main groups of CTs. CTs appearing in Table 5 have been arranged into two categories.

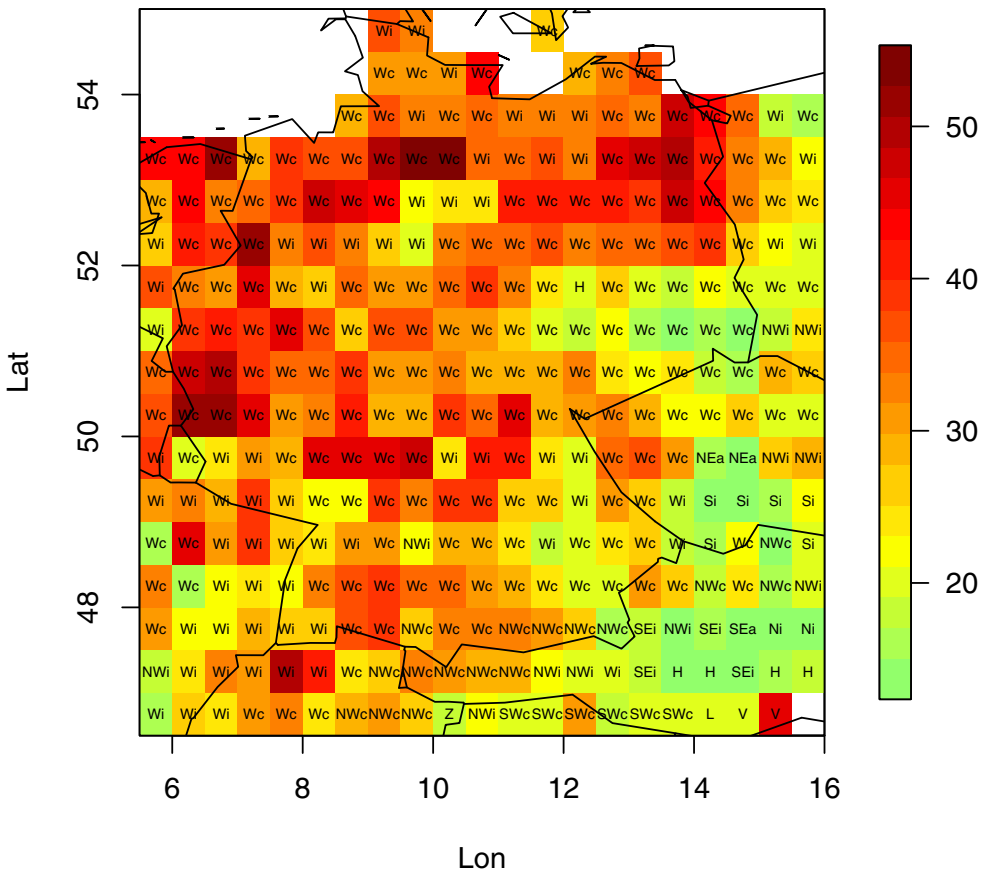

Fig. 10 Spatial distribution of the relative importance (explained variation) of the most important predictor of the best GWT-based multiple regression model at each grid cell for 3-month SPI in February. See text for details on the estimation of relative importance. Abbreviations in each grid cell indicate the respective circulation type or monthly circulation index

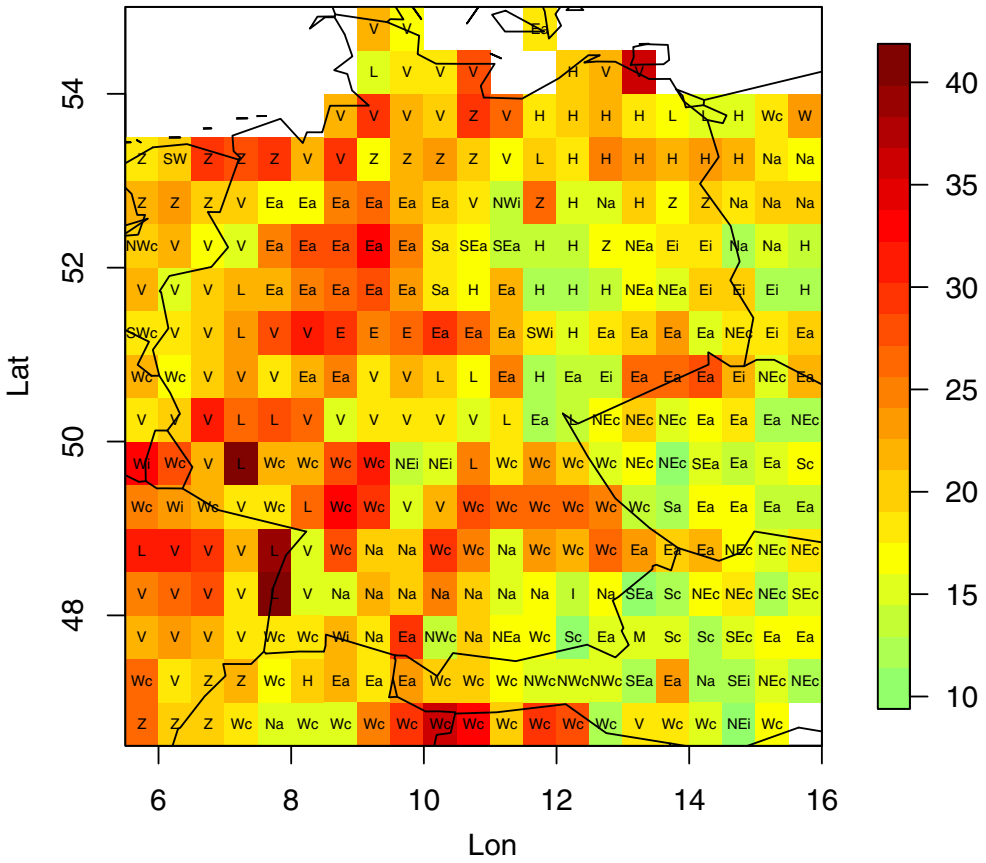

Fig. 11 Same as Fig. 10 but for May

The first category comprises CTs characterised by a southwesterly to northwesterly air mass advection or a central low/high pressure pattern (Wc, Wi, SWc, NWc, NWi, L, H). The second category includes CTs featuring advection from the easterly sector or pronounced meridional flow (NEc, NEa, Na, Ea, Si). February and November exhibit a clear preponderance of CTs from the first group, whereas appearance frequencies of both groups are far more balanced in May and August.

Mean relative importance (averaged over all grid cells) of predictors reach up to 33.0 % for the first predictor and up to 13.3 % for predictors 2 to 4. However, values of relative importance for individual grid cells may deviate distinctly from the overall mean values as can be seen—exemplary for the most important predictor in each month—from Figs. 10, 11, 12 and 13.

Concerning the most important predictor in February (Fig. 10), the clear preponderance of CTs Wc and Wi is obvious, especially in the northwestern, western and southwestern regions. Highest amounts of explained variance for Wc/Wi—partly exceeding 50 %—occur particularly in the northwestern parts of the Central European domain. On the other hand, mainly for grid cells in the east/southeast explained variances decrease to 20 % and even below.

In May *V*, Ea and Wc are the most frequent leading predictors. Maximum explained variances scarcely exceed 40 %, while minimum values are around 10 %. Although no clear-cut spatial pattern concerning the leading predictor becomes obvious, it appears that *V* is dominating in the western parts of the domain, whereas in the eastern parts no distinct preponderance of any predictor can be stated.

A much more distinct spatial structure is evident in August, with *V* being the predominant predictor in the west and northwest with explained variances up to 60 %. In the east/southeast, *V* becomes less frequent with explained variations declining to 10 %, while mainly H, Ea and NEa become more frequent. *V* is also in November the most frequent leading predictor featuring maximum/minimum values of explained variance around 45 % and 15 %, respectively. Wc, Wi and NWc reach comparable maximum/minimum values of relative importance. As in May, no distinct spatial patterns of preferred predictors are discernible in November.

4 Discussion and conclusions

The present investigations aimed for the quantification of the relationships between interannual variations in gridded 3-

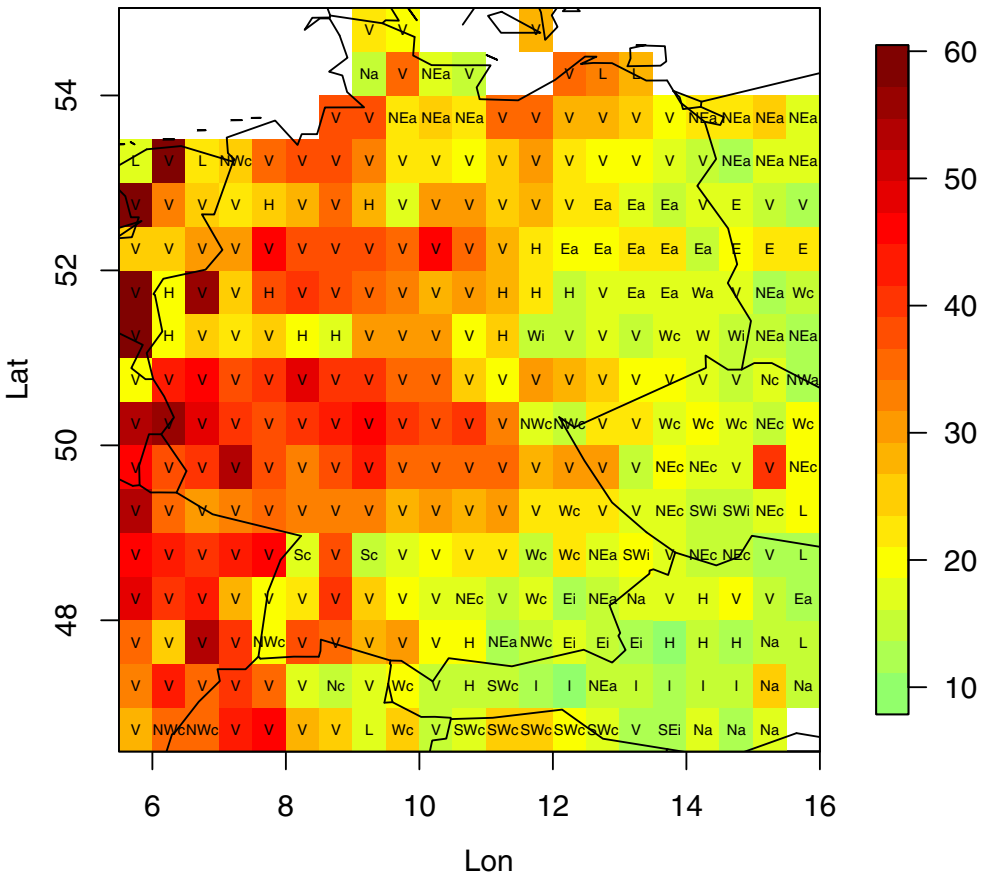

Fig. 12 Same as Fig. 10 but for August

month SPI over Central Europe and the large-scale atmospheric circulation. Therefore, we applied a variety of automated CTCs to daily gridded SLP data and utilized monthly occurrence frequencies of resulting types, together with monthly mean circulation indices, as predictors in multiple linear regression analyses to estimate corresponding SPI values for February, May, August and November. A resampling approach has been used to find adequate sets of predictors, and estimates for the relative importance of predictors have been derived. Based on a leave-one-out cross-validation procedure and the determination of several skill scores, the performance of all models has been assessed

For each grid cell, the 20 best performing models (approximately 10 % of all models created for each grid cell) according to the MSSS have been selected and further analysed concerning their skill characteristics and their configuration settings.

Some common features of the best performing models became evident.

Most of the superior models are based on the GWT classification, and most of the superior models utilize classifications comprising 18 or 27 CTs. The predominance of the GWT classification—particularly in February and November—can be explained by the fact that the GWT classification explicitly discriminates between cyclonic and anticyclonic CTs and thus—like other threshold-based CTCs—appears to be better suited for capturing precipitation variations than other CTCs (Beck and Philipp 2010). The clear preponderance of 18 and 27 types CTCs reflects the better separation of cyclonic and anticyclonic CTs with higher type numbers (e.g. the 10 types variant of the GWT classification does not incorporate the discrimination into cyclonic and anticyclonic subtypes).

With respect to the preferred properties of the spatial domain used for performing the CTCs, it becomes apparent that smaller domains are generally preferred in the majority of the best performing models. This, together with the fact that the predominance of the smaller domains appears most pronounced in August and November and less distinct in February and May, is in line with respective findings from Beck et al. (2013), reflecting seasonal variations concerning the spatial scale of circulation features that are relevant for precipitation.

Concerning the location of the domain on the other hand no clear-cut evidence for preferred settings shows up. This may be due to two interrelated reasons. Firstly, the variations in location of domains introduced here are rather small and therefore—in most cases—do not lead to the consideration of different synoptic systems. Secondly, the relevant characteristics of a specific circulation configuration may be captured focusing on different subparts of the related synoptic

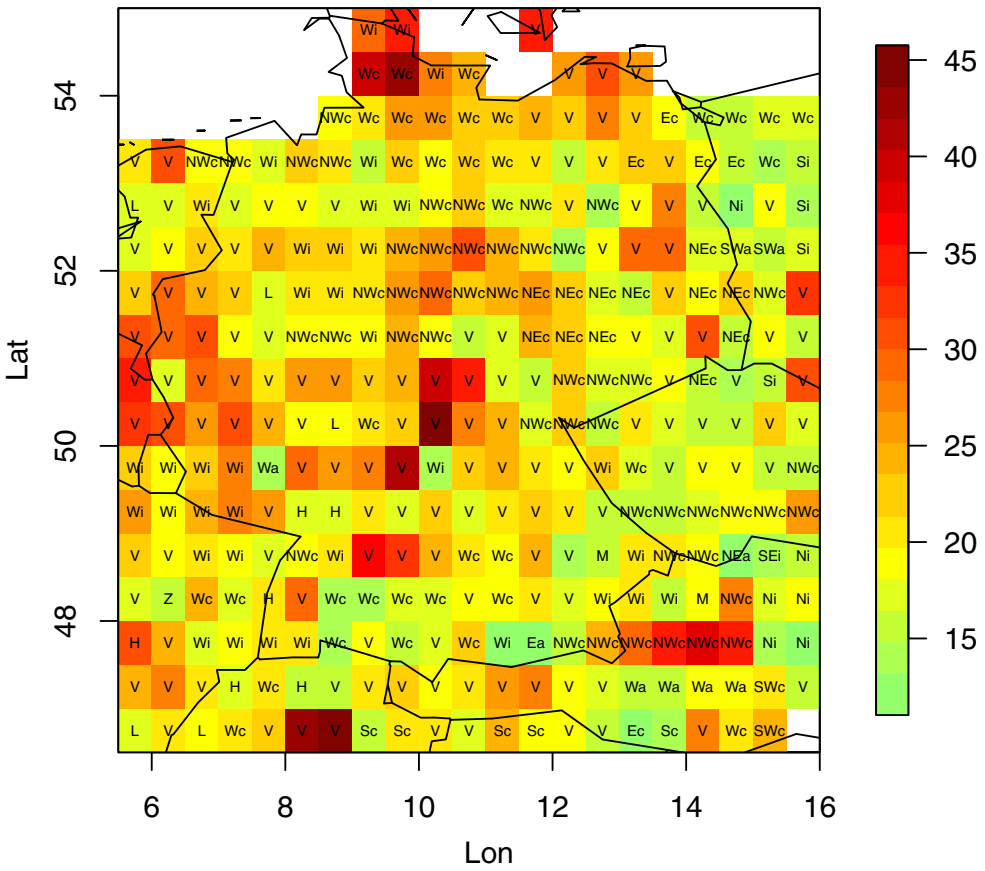

Fig. 13 Same as Fig. 10 but for November

systems—as has been shown by Beck et al. (2013). However, more systematic analyses should be performed to get a deeper insight into the relevance of the location of the classification domain.

The last configuration setting that has been varied is the consideration of monthly mean circulation indices (*V*, *M*, *Z*) as potential predictors in the MRAs in addition to CT frequencies. Apparently, a large majority of best performing models utilizes mean circulation indices as additional predictors. This can be seen most distinctly in August and is least obvious in February. Although the metrically scaled circulation indices used here do not reflect respective within-type variations of individual CTs (Beck et al. 2007), they obviously provide additional information on general circulation characteristics that are relevant for SPI variability. This is especially true for the most often selected *V* index reflecting the mean vorticity characteristics.

Although quite a large number of potential predictors (up to 30) are available, only a few predictors (2 to 5 in more than 80 % of all models) are selected for developing the final MRMs.

On the basis of the selected best performing models, distinct seasonal and spatial variations in skill become apparent. With respect to all skill scores, a generally better model performance in February than in August can be stated, while May and November exhibit intermediate performances. A more detailed picture of intra-annual variations in model skill could be achieved by analysing moving 3-month periods with monthly time steps. Such additional analyses exceed the scope of this paper but will be considered in future studies.

Spatial variations may be generalized as a decrease in model skill from the northwestern to the southeastern parts of the Central European domain. These findings are in accordance with respective results from other studies on the relationships between CTs and precipitation (Beck and Philipp 2010, Schiemann and Frei 2010, Beck et al. 2013). On the one hand, the detected variations reflect seasonal differences in the intensity of the coupling between large-scale circulation and precipitation being strongest in winter and weakest in summer, when smaller scale dynamic processes become increasingly important. On the other hand, such a gradient in the strength of the relationship between large-scale atmospheric dynamics and precipitation also exists between the western and northwestern—more oceanic—parts and the eastern and southeastern—more continental—parts of the Central European domain. At least partly, these seasonal and spatial variations are also discernible from the spatial patterns of CTs preferably selected as predictors. In the western parts of the domain a preponderance of CTs implying oceanic influence (southwesterly, westerly and northwesterly CTs) and of the *V* index

persists in all months reflecting the dominating synoptic-scale atmospheric features (e.g. mid-latitude cyclones and their frontal systems) relevant for precipitation in these regions. In regions farther east, CTs representing more continental influence (northerly to southerly circulation types) gain importance, particularly in May and August.

Finally, the important question in how far the presented approaches have the potential for downscaling of future SPI variability in Central Europe should be addressed, although they cannot be answered in a general way. The best models reach remarkable skill in terms of MSSS and as well in terms of HSSdry and HSSwet, particularly in the more oceanic, western parts of the domain. For these regions, featuring MSSS, HSSdry and HSSwet values well above 0.6, the utilization of the CTC-based MRMs within the framework of downscaling attempts for estimating possible future SPI variations appears to be feasible. However, for those regions farther to the east/southeast, exhibiting values of the varying skill scores close to 0 (indicating that the quality of the best models is comparable to an estimate based on climatology), the application of the respective CTC-based models in downscaling approaches is not justified.

It is worth mentioning that our analyses suggest that using the mean of the best performing models instead of the individual best model results to estimate SPI values leads to an improvement in model performance, in terms of all skill scores used in this study. However it needs to be further investigated in how far this finding is generally transferable.

For the present analyses, CTCs have been used that only consider gridded SLP data. However, recent studies (Huth et al. 2014) document that an improvement in synoptic skill of CTCs can be reached by the inclusion of additional large-scale atmospheric fields (e.g. vorticity, geopotential height of varying pressure levels, temperature) into the classification approaches. Such "extended" CTCs have been developed within the framework of the COST733 Action "Harmonisation and Applications of Weather Types Classifications for European Regions" (Philipp et al. 2014). Furthermore, CTCs may be further modified in order to attain maximum synoptic skill for specific target variables—e.g. gridded SPI—via variable weighting schemes applied to the variables entering the classification or by taking into consideration the target variable itself (so called conditional classifications).

The utilization of such optimized CTCs within MRMs as presented here may lead to a further improvement in model skill, possibly enabling the application of such approaches for the downscaling of future SPI variations, even in regions for which only modest skill has been reached in this contribution.

Acknowledgments Initial analyses contributing to the results presented in this paper have been performed within the COST733 Action "Harmonisation and Applications of Weather Types Classifications for European Regions" that has been funded by the EU within the framework of COST (European Cooperation in Science and Technology).

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723

Beck C, Jacobeit J, Jones PD (2007) Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in Central Europe since 1780. Int J Climatol 27:473–491

Beck C, Philipp A (2010) Evaluation and comparison of circulation type classifications for the European domain. Phys Chem Earth 35:374–387

Beck C, Philipp A, Streicher F (2013) The effect of domain size on the relationship between circulation type classifications and surface climate. Int J Climatol. doi:10.1002/joc.3688

Beck C, Weitnauer C, Jacobeit J (2014) Downscaling of monthly PM10 indices at different sites in Bavaria (Germany) based on circulation type classifications. Atmospheric Pollution Research. doi:10.5094/APR.2014.083

Bonsal BR, Aider R, Gachon P, Lapp S (2012) An assessment of Canadian prairie drought: past, present, and future. Clim Dyn. doi:10.1007/s00382-012-1422-0

Bordi I, Fraedrich K, Petitta M, Sutera A (2007) Extreme value analysis of wet and dry periods in Sicily. Theor Appl Climatol 87:61–71

Bordi I, Fraedrich K, Sutera A (2009) Observed drought and wetness trends in Europe: an update. Hydrol Earth Syst Sci 13:1519–1530

Bothe O, Fraedrich K, Zhu X (2010) The large-scale circulations and summer drought and wetness on the Tibetan plateau. Int J Climatol 30:844–855

Burke EJ, Brown SJ (2008) Evaluating uncertainties in the projection of future drought. J Hydrometeorol 9:292–299

Enke W, Spekat A (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression. Clim Res 8:195–207

Fink AH, Brücher T, Krüger A, Leckebusch GC, Pinto JG, Ulbrich U (2004) The 2003 European summer heatwaves and drought—synoptic diagnosis and impacts. Weather 59:209–216

Fleig AK, Tallaksen LM, Hisdal H, Stahl K, Hannah DM (2010) Inter-comparison of weather and circulation type classifications for hydrological drought development. Phys Chem Earth 35:507–515

Fleig AK, Tallaksen LM, Hisdal H, Hannah DM (2011) Regional hydrological drought in north-western Europe: linking a new Regional Drought Area Index with weather types. Hydrol Process. doi:10.1002/hyp.7644

Garcia-Herrera R, Paredes D, Trigo RM, Trigo IF, Hernandez E, Barriopedro D, Mendes AM (2007) The outstanding 2004/5 drought in the Iberian Peninsula: associated atmospheric circulation. J Hydrometeorol 8:483–498

Gerstengarbe FW, Werner PC, Rüge U (1999) Katalog der Grosswetterlagen Europas nach Paul Hess und Helmuth Brezowsky 1881–1992, 5th edn. Potsdam-Institut für Klimafolgenforschung (PIK): Potsdam, Germany

Grömping U (2007) Estimators of relative importance in linear regression based on variance decomposition. Am Stat 61:139–147

Hannaford J, Lloyd-Hughes B, Keef C, Parry S, Prudhomme C (2011) Examining the large-scale spatial coherence of European drought using regional indicators of precipitation and streamflow deficit. Hydrol Process 25:1146–1162

Hartigan JA (1975) Clustering algorithms. Wiley, New York

Hayes M, Svoboda M, Wall N, Widhalm M (2011) The Lincoln declaration on drought indices: universal meteorological drought index recommended. Bull Am Meteorol Soc 92:485–488

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded dataset of surface

temperature and precipitation. J Geophys Res Atmos. doi:10.1029/2008JD10201
Heim RR (2000) Drought indices: a review. In: Wilhite DA (ed) Drought: a global assessment, 1st edn. Routledge, London, pp 159–167
Heim RR (2002) A review of twentieth century drought indices used in the United States. Bull Am Meteorol Soc 83:1149–1165
Heinrich G, Gobiet A (2012) The future of dry and wet spells in Europe: a comprehensive study based on the ENSEMBLES regional climate models. Int J Climatol 32:1951–1970
Hess P, Brezowsky H (1952) Katalog der Grosswetterlagen Europas. Berichte des Deutschen Wetterdienstes in der US Zone 33. Deutscher Wetterdienst, Bad Kissingen, Germany
Huth R, Beck C, Philipp A, Demuzere M, Untstrul Z, Cahynova M, Kysely J, Tveito OE (2008) Classifications of atmospheric circulation patterns: recent advances and applications. Ann N Y Acad Sci 1146:105–152
Huth R (2010) Synoptic-climatological applicability of circulation classifications from the COST733 collection: First results. Phys Chem Earth 35:388–394
Huth R, Beck C, Tveito OE (2010) Classifications of atmospheric circulation patterns—theory and applications—preface. Phys Chem Earth 35:307–308
Huth R, Beck C, Cahynova M (2014) Synoptic-climatological evaluation of the classifications of atmospheric circulation patterns over Europe. submitted to Int. J. Climatol
James PM (2007) An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe. Theor Appl Climatol 88:17–42
Jenkinson AF, Collison FP (1977) An initial climatology of gales over the North sea. Synoptic Climatology Branch Memorandum 62. Meteorological Office, London
Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White J, Woolen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–470
Keyantash J, Dracup JA (2002) The quantification of drought: an evaluation of drought indices. Bull Am Meteorol Soc 83:1167–1180
Krämer W, Sonnberger H (1986) The linear regression model under test. Physica, Heidelberg
Lamb HH (1972) British Isles weather types and a register of daily sequence of circulation patterns, 1861 – 1971. Geophysical Memoir 116, HMSO, London
Lindeman RH, Merenda PF, Gold RZ (1980) Introduction to bivariate and multivariate analysis. Scott, Foresman, Glenview
Lloyd-Hughes B, Saunders MA (2002) A drought climatology for Europe. Int J Climatol 22:1571–1592
Lopez-Moreno JI, Vicente-Serrano SM (2008) Positive and negative phases of the Wintertime North-Atlantic Oscillation and drought occurrences over Europe: a multitemporal-scale approach. J Clim 21:1220–1242
Loukas A, Vasiliades L, Tzabiras J (2007) Evaluation of climate change on drought impulses in Thessaly, Greece. Eur Water 17(18):17–28
McKee TBN, Doesken J, Kleist J (1993) The relationship of drought frequency and duration to time scales. Eight Conference On Applied Climatology. American Meteorological Society, Anaheim, CA, pp 179–184
McKee TBN, Doesken J, Kleist J (1995) Drought monitoring with multiple time scales. Ninth Conference On Applied Climatology. American Meteorological Society, Dallas, TX, pp 233–236
Murphy AH (1988) Skill scores based on the mean square error and their relationship to the correlation coefficient. Mon Weather Rev 116: 2417–2424
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, Pianko-Kluczynska K, Post P, Rassilla Álvarez D, Schiemann R, Spekat A, Tymvios FS (2010) COST733CAT—a database of weather and circulation type classifications. Phys Chem Earth 35:360–373
Philipp A, Beck C, Huth R, Jacobeit J (2014) Development and comparison of circulation type classifications using the COST 733 dataset and software. Int J Climatol. doi:10.1002/joc.3920
Pongracz R, Bogardi I, Duckstein L (1999) Application of fuzzy rule-based modeling technique to regional drought. J Hydrol 224:100–114
Pongracz R, Bogardi I, Duckstein L (2003) Climatic forcing of droughts: a Central European example. Hydrol Sci 48:39–50
Raziei T, Bordi I, Pereira LS, Corte-Real J, Santos JA (2012) Relationship between daily atmospheric circulation patterns and winter dry/wet spells in western Iran. Int J Climatol 32:1056–1068
Royston P (1982) An extension of Shapiro and Wilk's W test for normality to large samples. Appl Stat 31:115–124
Santos JF, Pulido Calvo I, Portela MM (2010) Spatial and temporal variability of droughts in Portugal. Water Resour Res 46:W03503
Schiemann R, Frei C (2010) How to quantify the resolution of surface climate by circulation types: an example for Alpine precipitation. Phys Chem Earth 35:403–410
Sienz F, Bordi I, Fraedrich K, Schneidereit A (2007) Extreme dry and wet events in Iceland: observations, simulations and scenarios. Meteorol Z 16:9–16
Sienz F, Bothe O, Fraedrich K (2012) Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias. Hydrol Earth Syst Sci 16:2143–2157
Stahl K, Demuth S (1999) Linking streamflow drought to the occurrence of atmospheric circulation patterns. Hydrol Sci J 44:467–482
Tran L, Knight CG, Wesner V (2002) Drought in Bulgaria and atmospheric synoptic conditions over Europe. Geol J 57:149–157
Trnka M, Kysely J, Mozny M, Dubrovsky M (2009) Changes in Central European soil moisture availability and circulation patterns in 1881–2005. Int J Climatol 29:655–672
Vasiliades L, Loukas A, Patsonas G (2009) Evaluation of a statistical downscaling procedure for the estimation of climate change impacts on droughts. Nat Hazard Earth Syst 9:879–894
Vicente-Serrano SM, Lopez-Moreno JI (2006) The influence of atmospheric circulation at different spatial scales on winter drought variability through a semi-arid climatic gradient in northeast Spain. Int J Climatol 26:1427–1453
Wilks DS (2006) Statistical methods in atmospheric sciences, 2nd edn. Elsevier, New York