Earth Syst. Dynam., 8, 951–962, 2017
https://doi.org/10.5194/esd-8-951-2017

Earth System Dynamics
Open Access
EGU

Climate indices for the Baltic states from principal component analysis

Liga Bethere, Juris Sennikovs, and Uldis Bethers

Laboratory for Mathematical Modelling of Environmental and Technological Processes, University of Latvia, Riga LV-1002, Latvia

Correspondence to: Liga Bethere (liga.bethere@lu.lv)

Received: 30 March 2017 – Discussion started: 6 April 2017
Revised: 5 September 2017 – Accepted: 19 September 2017 – Published: 26 October 2017

Abstract. We used principal component analysis (PCA) to derive climate indices that describe the main spatial features of the climate in the Baltic states (Estonia, Latvia, and Lithuania). Monthly mean temperature and total precipitation values derived from the ensemble of bias-corrected regional climate models (RCMs) were used. Principal components were derived for the years 1961–1990. The first three components describe 92 % of the variance in the initial data and were chosen as climate indices in further analysis. Spatial patterns of these indices and their correlation with the initial variables were analyzed, and it was detected (based on correlation coefficient between principal components and initial variables) that higher values in each index corresponded to locations with (1) less distinct seasonality, (2) warmer climate, and (3) wetter climate. In addition, for the pattern of the first index, the impact of the Baltic Sea (distance to coast) was apparent; for the second, latitude and elevation were apparent, and for the third elevation was apparent. The loadings from the chosen principal components were further used to calculate the values of the climate indices for the years 2071–2100. An overall increase was found for all three indices with minimal changes in their spatial pattern.

1 Introduction

Spatial representation of the climate, e.g., the mapping of climatic zones, is a useful tool in climate analysis. First, it can be used to better convey information about the climate features of the region for applications in climate change adaptation and mitigation. Second, the spatial patterns can give insight into both the possible relationship between and the impact of the climate on other fields, e.g., phenological processes and vegetation distribution (Feng et al., 2012). Third, they illustrate geographical features that influence climate, such as hillsides and coastal zones. There is a wide variety of approaches for creating spatial representations of climate, but usually they belong to either rule-driven or data-driven methods. Rule-driven methods are used more often, the most popular being the Köppen–Geiger classification (Peel et al., 2007). These methods are based on certain predefined rules; for example, thresholds of meteorological variables or frequency of events. Climate zones derived from classifications of this type usually correspond to vegetation distributions in the sense that each climate type is dominated by one vegetation zone or eco-region (Belda et al., 2014). However, predefined rules make these methods subjective. Alternatively, the spatial pattern can be derived from data-driven or analytical methods. These include principal component analysis (PCA; Benzi et al., 1997; Estrada et al., 2009), cluster analysis (Bieniek et al., 2012), or a combination of both methods (Briggs and Lemin, 1992; Fovell and Fovell, 1993; Baeriswyl and Rebetez, 1997; Malmgren et al., 1999; Fan et al., 2014; Forsythe et al., 2015). Analytical methods, depending on the chosen variables, can give results that are similar to those of rule-driven methods, but the results are more homogenous (Netzel and Stepinski, 2016). Analytical methods provide a spatial pattern that must be interpreted before it can be linked with possible applications.

Principal component analysis or empirical orthogonal function analysis has two important applications. First, it can reduce the number of variables that are used to describe regional climate while still retaining most of the variation seen

in the initial data. Second, principal components provide new indices that are a linear combination of the chosen variables. The loadings of the chosen principal components are the coefficients that define the newly created indices, which then describe the main features of climate. Variables for PCA can be chosen and indices calculated with a specific purpose in mind; for example, indices for the classification of different types of winters (Hagen and Feistel, 2005) or estimation of crop yield based on the climate (Cai et al., 2013). Indices can also be chosen to describe the climate of the region in general (Estrada et al., 2009). However, the problem with the indices that are derived using analytical methods is that their meaning is not known beforehand, so their interpretation may require further analysis.

For many practical applications, temperature and precipitation are the two main variables of interest for a certain region. They are usually sufficient for representing vegetation types in corresponding climate zones (Zhang and Yan, 2014). Vegetative production, organic matter decomposition, and the cycling of nutrients are strongly influenced by temperature and moisture (Briggs and Lemin, 1992). Distinct changes in temperature and precipitation are to be expected in the future (BACC II, 2015). Thus, any climate patterns based on these two variables will consequently be affected, leaving a significant impact on living organisms. For instance, plant species inhabiting regions subjected to climate change might have too little time to adapt (Mahlstein et al., 2013).

The Baltic state region exhibits significant spatial and temporal climatic variability, with an influence from air masses of arctic to subtropical origin (Jaagus and Ahas, 2000; Rutgersson et al., 2014). The terrain is mostly flat, with the highest elevations extending slightly above 300 m. The Baltic Sea and the shape of its coastline have an important role in the climate of the region. PCA has been used to describe precipitation patterns in the Baltic countries with atmospheric and landscape variables (Jaagus et al., 2010).

To study the effects of climate change on climate patterns, regional climate model (RCM) data can be used (Castro et al., 2007; Mahlstein and Knutti, 2010; Tapiador et al., 2011; Fan et al., 2014). RCMs are continuously improving and correspond rather well to climate observations (Tapiador et al., 2011). Other advantages of using RCM data are that (a) their data are regularly spaced, while PCA applied to irregularly spaced data can produce distorted loading patterns (Karl et al., 1982), and (b) RCM data are also available as future projections, giving insight into the manifestation of climate change. Additionally, the spatial representativeness of the network of observation stations in the Baltic states has been reported to be problematic (Remm and Jaagus, 2011).

The aim of this work is to define climate indices that represent the main features of Baltic state climate in a compact form. The study consists of several parts. First, RCM data for temperature and precipitation were bias corrected. Second, monthly average values for the reference period 1961–1990 were calculated and standardized. Third, PCA was performed and the main principal components were identified. The acquired principal components and their spatial patterns were analyzed. Fourth, the loadings of chosen principal components were used to calculate indices for the years 2071–2100 and compared to reference data.

Table 1. List of the regional climate model (RCM) ensemble members used (ENSEMBLES) showing the originating institution, the name of the RCM, and the driving general circulation model (GCM). For an explanation of abbreviations, see van der Linden and Mitchell (2009).

Institution	GCM	RCM
C4I	HadCM3Q16	RCA3
CNRM	ARPEGE	Aladin
CNRM	ARPEGE_RM 5.1	Aladin
DMI	ARPEGE	HIRHAM
DMI	ECHAM5-r3	DMI-HIRHAM5
ETHZ	HadCM3Q0	CLM
GKSS	IPSL	CLM
HC	HadCM3Q0	HadRM3Q0
HC	HadCM3Q16	HadRM3Q16 (high sensitivity)
HC	HadCM3Q3	HadRM3Q3 (low sensitivity)
ICTP	ECHAM5-r3	RegCM
KNMI	ECHAM5-r3	RACMO
KNMI	ECHAM5-r3	RACMO
KNMI	MIROC	RACMO
METNO	BCM	HIRHAM
METNO	HadCM3Q0	HIRHAM
MPI	ECHAM5-r3	REMO
SMHI	BCM	RCA
SMHI	ECHAM5-r3	RCA
SMHI	HadCM3Q3	RCA
UCLM	HadCM3Q0	PROMES
VMGO	HadCM3Q0	RRCM

2 Data and methods

2.1 Climate data and methods

The source of the RCM ensemble data is the ENSEMBLES project (van der Linden and Mitchell, 2009). Model data sets for the A1B scenario are given for the time period 1961–2100, and 22 model runs were considered (shown in Table 1).

We used time series of daily average air temperature at 2 m of height and daily precipitation. RCMs are known to be prone to systematic biases (Teutschbein and Seibert, 2012). A bias correction method (Sennikovs and Bethers, 2009) that uses quantile mapping was chosen and the cumulative distribution function was calculated for each day of the year using an 11-day running average – the data for 5 days before and 5 days after the day of interest. The ensemble median was then used for PCA. The control period for bias correction

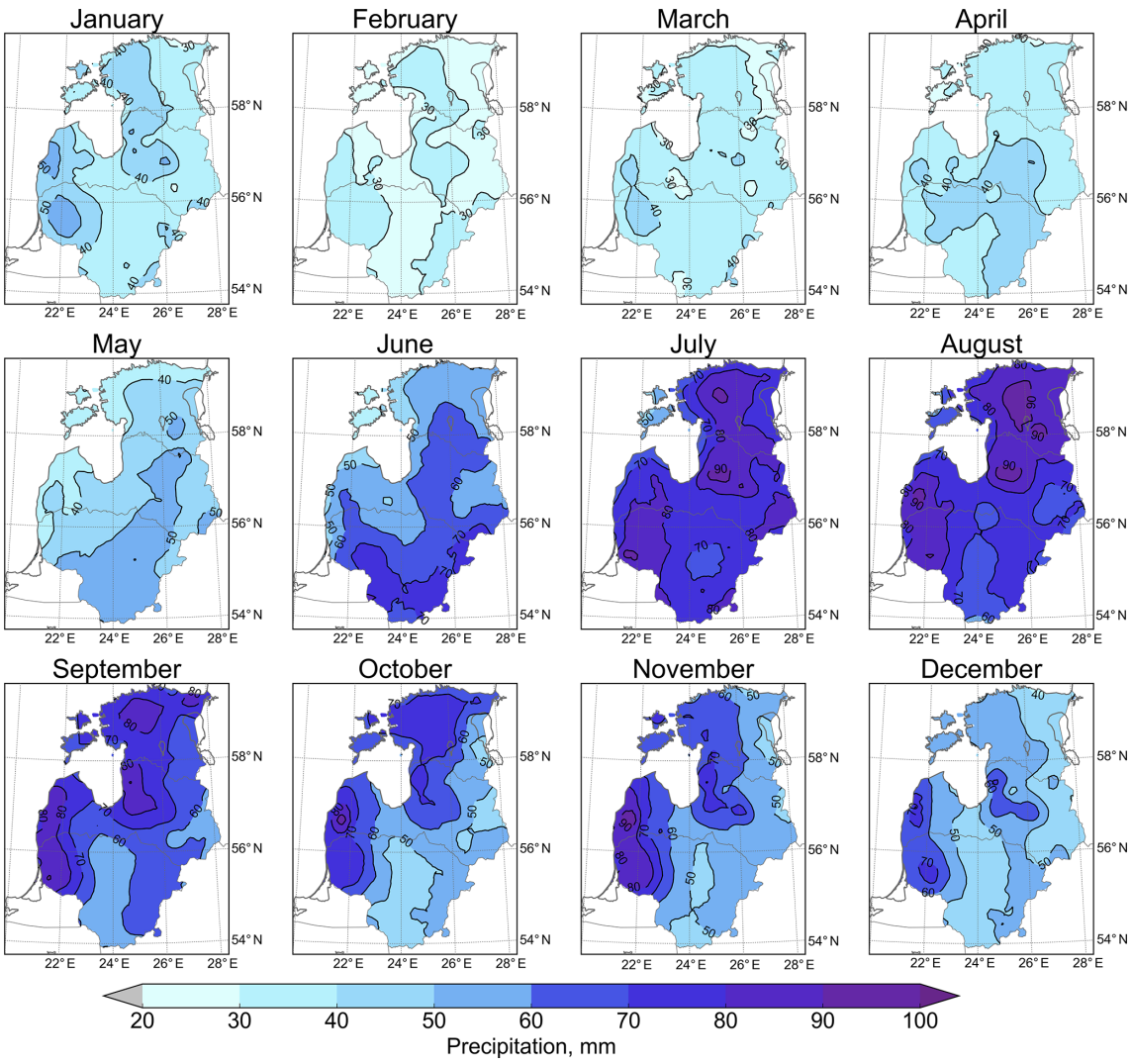

Figure 1. Monthly precipitation 1961–1990; bias-corrected median of RCM ensemble.

was 1961–1990. Bias-corrected data were then interpolated to a regular grid because it has been shown that PCA applied to irregularly spaced data can produce distorted loading patterns (Karl et al., 1982). The bias correction method and model resolution is described in detail in Sennikovs and Bethers (2009).

Two time periods were chosen: 1961–1990 (as a reference climate) and 2071–2100 (as future climate projections). For each time period, monthly average temperature and precipitation were calculated for each grid point. In total 24 climatic variables were used for each time period: 12 monthly precipitation and 12 monthly average temperatures. This is an "R-mode" analysis according to Cattell (1952). The spatial distribution of these variables for the reference period is shown in Figs. 1 and 2. Figure 1 shows a north–south gradient of monthly precipitation during April–June and an east–west gradient of monthly precipitation during October–January. Figure 2 shows an east–west gradient of monthly temperatures during October–February and a north–south gradient of monthly temperatures during April–June. This implies that some of the variables can be combined in seasons (as done by Malmgren et al., 1999, and Forsythe et al., 2015) and that for some months temperature and precipitation are correlated. A better understanding of variables with similar patterns can be gained by examining the correlation matrix in Fig. 3. The matrix areas that represent strongly correlated variables are marked in this figure, and they show the following relationships.

1. *Very strong correlation (above 0.8) between precipitation levels in winter months*. Locations with more precipitation in, e.g., December also have more precipitation in January (compared to the rest of the territory).

2. *Strong correlation (above 0.5) between precipitation and temperature in spring months*. Thus, locations with

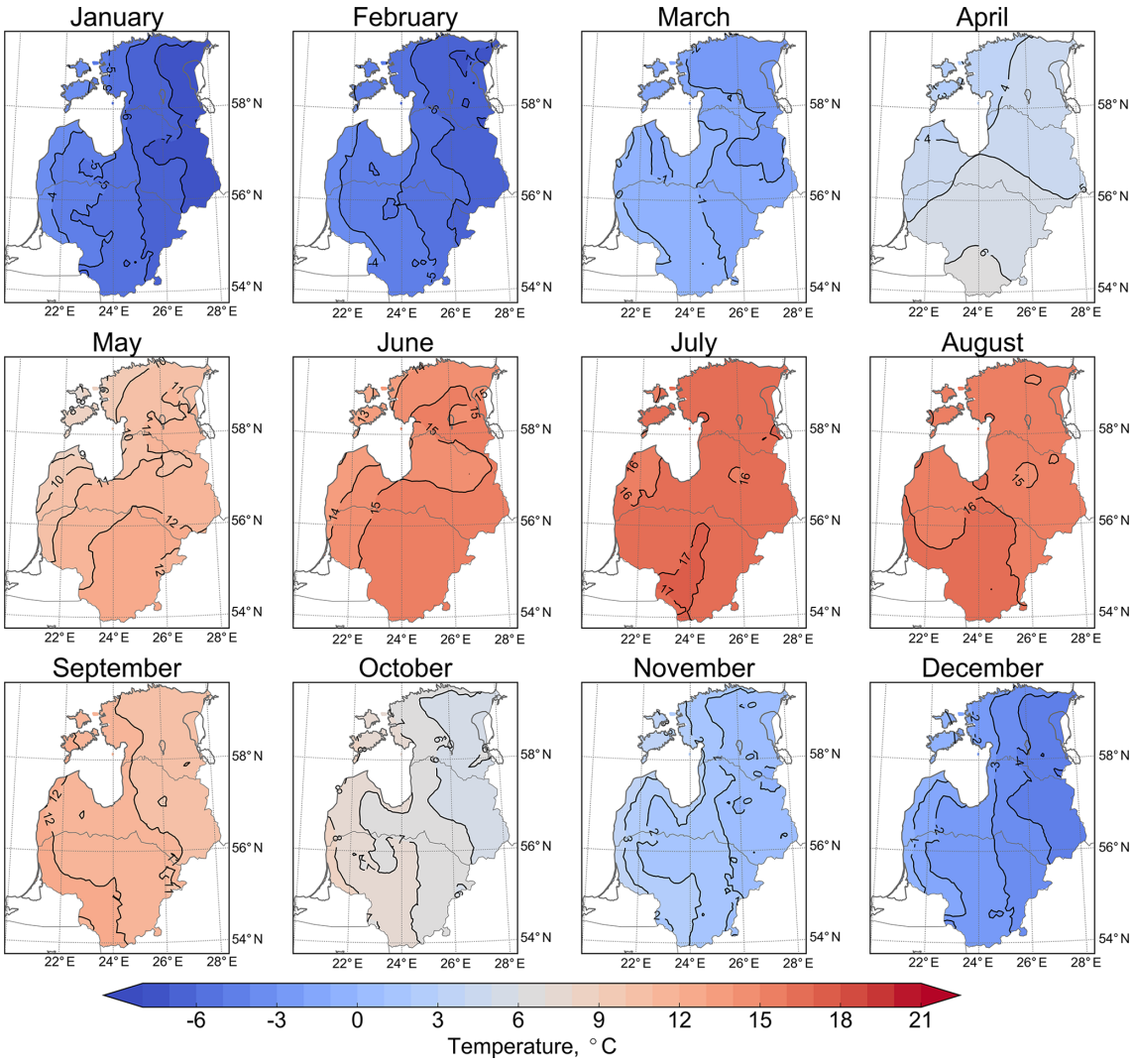

Figure 2. Monthly average temperature 1961–1990; bias-corrected median of RCM ensemble.

colder springs also are dryer, whilst locations with warmer springs also have more spring precipitation.

3. *Strong negative correlation (below −0.5) between precipitation in autumn and late spring/early summer temperature*. Locations with more precipitation in autumn also have colder springs.

4. *Very strong correlation (above 0.8) between temperatures of autumn and winter months*. Locations with warmer autumns also have warmer winters.

Figure 3 shows that the 24 monthly variables contain redundant information, and through PCA we can summarize the information and create new variables.

2.2 PCA method

The aim of PCA is to create a new set of uncorrelated variables that are a linear combination of the initial variables and explain as much of the initial variation as possible. An extensive description of PCA can be found in Jolliffe (2002), and its applications to climate are described in Preisendorfer (1988).

Although PCA is a widely used methodology, the terminology in the literature can vary (Wilks, 2011). We will briefly describe the terminology used in this article.

Suppose that $\mathbf{X}$ is an $n \times p$ data matrix, where n is the number of objects and p is the number of variables. The means of the p variables have been subtracted. In our case we have $p = 24$ climatic variables in $n = 7143$ grid points. A typical PCA is applied to $p \times p$ covariance (or correlation) matrix calculated by Eq. (1). By solving Eq. (2) we can find eigenvectors $\boldsymbol{e}_i$, $i = 1, \ldots, 24$ and corresponding eigenvalues

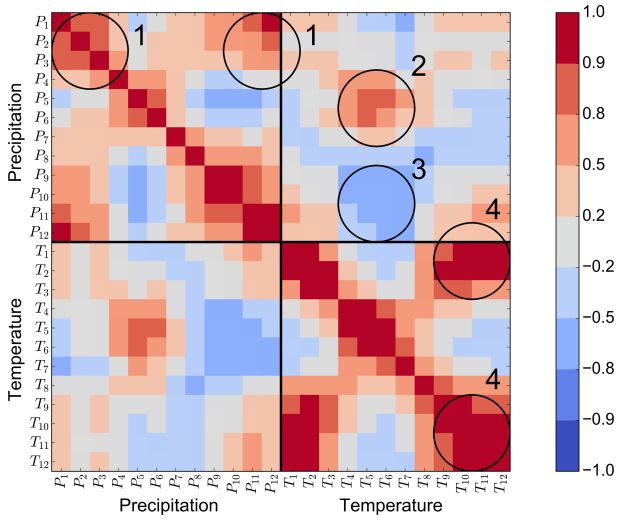

Figure 3. Temperature–precipitation correlation matrix; bias-corrected data. Marked and numbered features show especially high absolute correlation: (1) strong correlation between precipitation levels in winter months; (2) strong correlation between precipitation and temperature in spring months; (3) strong negative correlation between precipitation in autumn and spring temperature; (4) strong correlation between temperatures in autumn and winter months.

λ_i, $i = 1, \ldots, 24$. As a result we have obtained non-correlated linear combinations of the initial climatic variables calculated by Eq. (3).

$$\mathbf{S} = (n-1)^{-1}\mathbf{X}^T\mathbf{X} \quad (1)$$

$$\mathbf{S}\boldsymbol{e} = \lambda\boldsymbol{e} \quad (2)$$

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{e}_i \quad i = 1, \ldots, 24 \quad (3)$$

Values λ_i represent the explained variance of each "principal component" $\boldsymbol{Y}_i$. Linear weights $\boldsymbol{e}_i$ that define each principal component will be called "loadings". "Indices" describe $\boldsymbol{Y}_i$ values that are calculated using loadings from the reference period (but not necessarily reference period data). For the reference period, principal components coincide with indices, but indices can be also calculated using future period data and reference period loadings.

An important choice must be made when applying PCA: whether to use a correlation matrix or covariance matrix in the calculation of loadings. If the covariance matrix is used then a second choice must be made: whether to use standardization and what type. The scaling process has a significant impact on the PCA process. When performing data standardization, the following issues should be taken into account.

1. Variables should be of a similar scale; otherwise, variables with considerably larger variance will dominate the principal components. Different scales are usually a consequence of different units of measurement. In our case the variance for precipitation measured in millimeters is considerably larger than that for temperature measured in degrees Celsius.

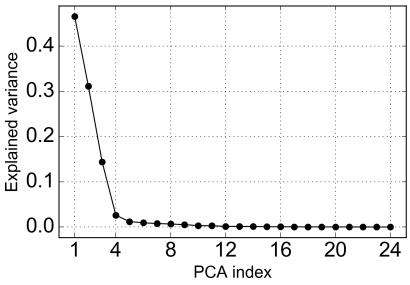

Figure 4. Scree plot (explained variance of each principal component) calculated for the reference (1961–1990) climate.

2. In the case of variables measured in the same units, variances contain useful information and can improve the interpretation of PCA (Overland and Preisendorfer, 1982). Therefore, for variables that are measured in the same units (for example, average temperature in different months) we wish to keep the ratio between variances of different months. This means that the correlation matrix, in which each variable is divided by its square root of variance, should not be used as it would bring the variances of all 24 variables to 1.

3. As we are planning to use the acquired loadings as coefficients for the calculation of climate indices for the future time period and compare them with the reference climate, it is necessary that the same standardization process be used for the data of the future time period.

4. It is important to note that subtraction of the mean (or a similar constant) for each variable does not impact the result of PCA as it does not impact the covariance between variables. However, if the initial values have a zero mean (the mean is subtracted from each variable) then the resulting principal components have a similar scale, and spatial patterns are more convenient to review.

Taking into account the issues described above we propose using standardization as defined by Eq. (4), in which the spatial mean is subtracted for each variable as usual, but the average variance of all temperature or precipitation variables is used for scaling:

$$\frac{\boldsymbol{T}_k - \overline{T}_k}{\sqrt{\overline{V}(T)}}, \quad \frac{\boldsymbol{P}_k - \overline{P}_k}{\sqrt{\overline{V}(P)}}, \quad k = 1, \ldots, 12, \quad (4)$$

where $\overline{V}(T)$, $\overline{V}(P)$ represents the average variance of 12 temperature and precipitation variables for the reference period.

The variances before and after such standardization for the reference period are shown in Table 2. The ratio of variances for different months is retained. For data representing the future time period, the standardization is performed by using the mean values and average variances from the reference

Table 2. Variances of climate variables before and after standardization for the years 1961–1990.

1961–1990												
Before standardization												
P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	Mean
28.85	7.45	13.03	13.66	31.93	63.40	47.20	65.65	86.22	110.43	114.47	50.60	52.74
T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	Mean
1.36	0.95	0.60	0.62	0.93	0.41	0.09	0.19	0.39	0.54	0.83	1.27	0.68
After standardization												
P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	Mean
0.55	0.14	0.25	0.26	0.61	1.20	0.89	1.24	1.63	2.09	2.17	0.96	1.00
T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	Mean
2.00	1.40	0.88	0.91	1.37	0.60	0.14	0.27	0.57	0.80	1.22	1.86	1.00

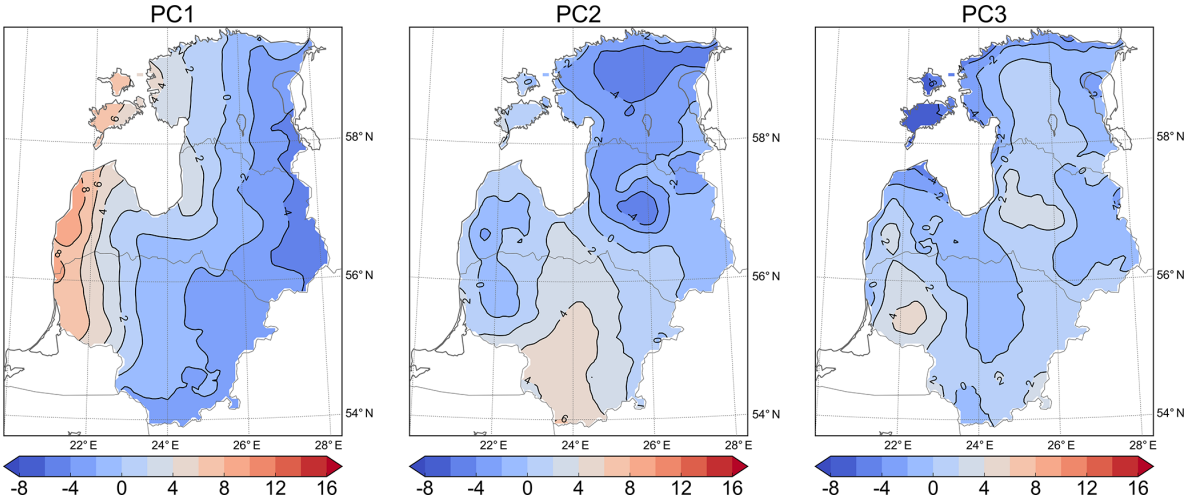

Figure 5. Spatial pattern of first three principal components based on monthly temperature and precipitation data for the years 1961–1990.

period. The results of data standardization for the future time period are shown in Table 3. It can be seen that in the future the variance in precipitation data will increase and the variance in temperature data will decrease. However, the distribution of variances over the year is similar.

Another detail that must be considered when using PCA is the choice of method for determining the number of principal components that describe data variation sufficiently well and can be used in further analysis. There are multiple methods to choose from (Preisendorfer, 1988); however, in our case one of the most common methods, the scree plot, gives excellent and clear results. A scree plot is a graph of explained variances in acquired principal components, and the number of principal components is decided based on the break point in such a graph. Components to the left of the break point are retained.

3 Results

3.1 Principal components for the control period (1961–1990)

The explained variance and loadings of the first three principal components are shown in Table 4. The scree plot of all principal components is shown in Fig. 4. The first two components already describe 78 % of the variance in the initial variables, while the first three components describe 92 % of the variance. According to Jolliffe (2002) the cutoff point should be between 70 and 90 % of the explained variance. However, the scree plot clearly shows that the first three principal components can be retained, so we chose to further analyze the first three components.

Figure 5 shows the spatial pattern of the first three principal components for the reference climate. They should be analyzed together with the correlation coefficients between the new variables and initial variables shown in Table 5, in

Table 3. Variances of climate variables before and after standardization for the years 2071–2100.

2071–2100												
Before standardization												
P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	Mean
52.78	12.33	22.68	27.02	33.84	52.5	42.87	72.7	126.1	154.3	204.3	85.6	73.92
T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	Mean
1.08	0.92	0.37	0.25	0.26	0.12	0.11	0.2	0.45	0.51	0.84	1.08	0.52
After standardization												
P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	Mean
1.00	0.23	0.43	0.51	0.64	1.00	0.81	1.38	2.39	2.93	3.87	1.62	1.40
T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	Mean
1.59	1.35	0.55	0.36	0.38	0.18	0.16	0.3	0.67	0.74	1.23	1.58	0.76

which the bright red or blue colors mark high positive or negative correlation. One can see that variables that were initially highly correlated (positively or negatively; Fig. 3) show similar (or in the case of negative correlation, the opposite) values in Table 5.

Correlation coefficient values (Table 5) show that the first principal component (PC1) has a high positive correlation with the autumn–winter temperature and precipitation and a high negative correlation with temperature and precipitation in late spring and early summer months. This means that higher values of PC1 correspond to places with warmer winters with more precipitation (snow or rain) and colder summers with less precipitation. However, it is also important to note that the total sum of the loadings is above 1, which implies that a constant increase in all variables would also result in higher values of PC1. From the spatial distribution (Fig. 5) we can see that PC1 has an east–west gradient implying less distinction between seasons at the seaside. It can be concluded that PC1 reflects the continentality of climate, and it represents the influence of the Baltic Sea.

The second principal component (PC2) is positively correlated with all monthly temperatures and negatively correlated with precipitation in autumn. This means that high PC2 values correspond to regions that are generally warmer than others and have low precipitation in autumn. For PC2 a north–south gradient is evident with the warmer climate in the south. This means that PC2 represents the influence of latitude. This pattern is also slightly influenced by geographical features (elevation) and the shape of the coast.

PC3 is mainly positively correlated with precipitation for most of the year (December–August) and spring temperature (April–May). This means that high PC3 values correspond to places with overall high precipitation or, in other words, an overall wetter year. PC3 mainly reflects the terrain, i.e., the distribution of elevation.

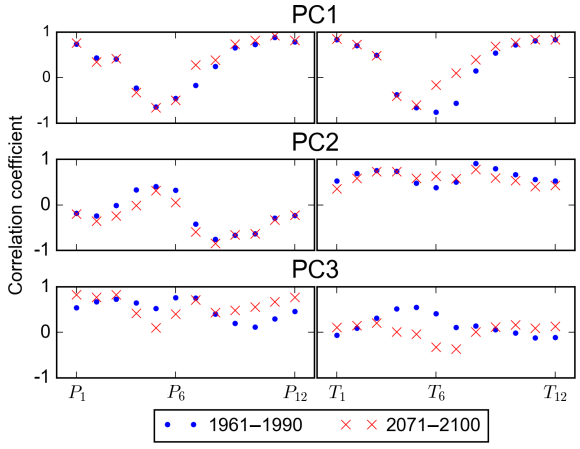

Figure 6. Correlation coefficients between indices (principal components) and initial variables for the reference and future climates.

When the spatial patterns of PC2 and PC3 are analyzed the effect of orography can be seen. The location of the highlands is especially visible, while for PC1 the terrain seems to have little impact.

3.2 Climate indices for future climate (2071–2100)

Loadings (linear weights) acquired through PCA from the reference data (Table 4) can be used as coefficients that define new climate indices. We can use these coefficients to calculate climate from different data (other time periods or other geographical locations). It is also important to note that statistics (mean values and variances) from the reference data used in data standardization should also be applied to other data for comparison to be possible. In our case we calculated such climate indices for future climate (corresponding to the period 2071–2100) and analyzed the change in climate patterns. The standardization of the variables is shown by Eq. (5), and the calculation of the climate indices is shown

Table 4. Explained variance and loadings of the first three principal components calculated from temperature and precipitation data for the years 1961–1990.

	PC1	PC2	PC3	Sum
Explained variance	0.47	0.31	0.14	0.92
Loadings				
P_1	0.16	−0.05	0.22	
P_2	0.05	−0.03	0.14	
P_3	0.06	0.00	0.20	
P_4	−0.03	0.06	0.18	
P_5	−0.15	0.12	0.22	
P_6	−0.15	0.13	0.45	
P_7	−0.05	−0.15	0.38	
P_8	0.08	−0.31	0.24	
P_9	0.25	−0.31	0.13	
P_{10}	0.32	−0.33	0.09	
P_{11}	0.39	−0.16	0.24	
P_{12}	0.23	−0.08	0.24	
T_1	0.35	0.27	−0.04	
T_2	0.25	0.30	0.06	
T_3	0.14	0.26	0.16	
T_4	−0.11	0.26	0.27	
T_5	−0.23	0.21	0.35	
T_6	−0.18	0.11	0.17	
T_7	−0.06	0.07	0.02	
T_8	0.02	0.17	0.04	
T_9	0.12	0.22	0.02	
T_{10}	0.19	0.22	−0.01	
T_{11}	0.27	0.23	−0.07	
T_{12}	0.34	0.27	−0.08	

by Eq. (6):

$$\frac{\boldsymbol{T}_k - \overline{T}_k}{\sqrt{\overline{V}(T)}}, \quad \frac{\boldsymbol{P}_k - \overline{P}_k}{\sqrt{\overline{V}(P)}}, \quad k = 1, \ldots, 12, \tag{5}$$

where $\boldsymbol{T}_k$, $\boldsymbol{P}_k$ represents temperature and precipitation values for the future period, $\overline{T}_k$, $\overline{P}_k$ represents mean temperature and precipitation values for the reference period, and $\overline{V}(T)$, $\overline{V}(P)$ represents the average variance in 12 temperature and precipitation variables for the reference period.

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{c}_i, \; i = 1, \ldots, 24, \tag{6}$$

where $\boldsymbol{X}_i$ represents temperature and precipitation data for the future period, $\boldsymbol{c}_i$ represents coefficients (loadings) from the reference period, and $\boldsymbol{Y}_i$ represents climate indices for the future period.

It is important to note that $\boldsymbol{Y}_i$ values should not be called "principal components" even though they hold a similar meaning as principal components from the reference data. $\boldsymbol{Y}_i$ values are not derived using PCA directly and they do not use eigenvectors from future data.

In Fig. 6 the correlation coefficients between indices and initial variables are shown and it can be seen that they are similar to those for past climate. Therefore, they have the same interpretation and it is possible to analyze the change in spatial patterns between the past and future climate. The spatial distributions of future indices are shown in Fig. 7. Statistical descriptors, e.g., the minimal, maximal, and mean value of past and future indices, are summarized in Table 6. In addition, as we have used the same standardization (subtraction of the reference period mean) and climate index calculation process (loadings from the reference period), we can derive conclusions about increases or decreases in these climate indices. However, it is important to note that no conclusions can be derived about the value by which the increase or decrease has happened.

Table 5. Correlation coefficients between principal components and standardized initial data for the years 1961–1990. High positive correlation corresponds to darker red color and high negative correlation corresponds to darker blue color.

	PC1	PC2	PC3
P_1	0.73	-0.18	0.54
P_2	0.44	-0.24	0.68
P_3	0.41	-0.01	0.73
P_4	-0.22	0.33	0.65
P_5	-0.65	0.4	0.53
P_6	-0.45	0.33	0.76
P_7	-0.17	-0.42	0.75
P_8	0.25	-0.75	0.41
P_9	0.66	-0.67	0.2
P_{10}	0.73	-0.63	0.12
P_{11}	0.89	-0.29	0.3
P_{12}	0.78	-0.23	0.46
T_1	0.83	0.53	-0.06
T_2	0.7	0.69	0.1
T_3	0.49	0.76	0.32
T_4	-0.38	0.74	0.52
T_5	-0.66	0.48	0.55
T_6	-0.76	0.38	0.41
T_7	-0.57	0.5	0.11
T_8	0.15	0.91	0.14
T_9	0.54	0.8	0.06
T_{10}	0.72	0.67	-0.01
T_{11}	0.81	0.56	-0.12
T_{12}	0.83	0.53	-0.11

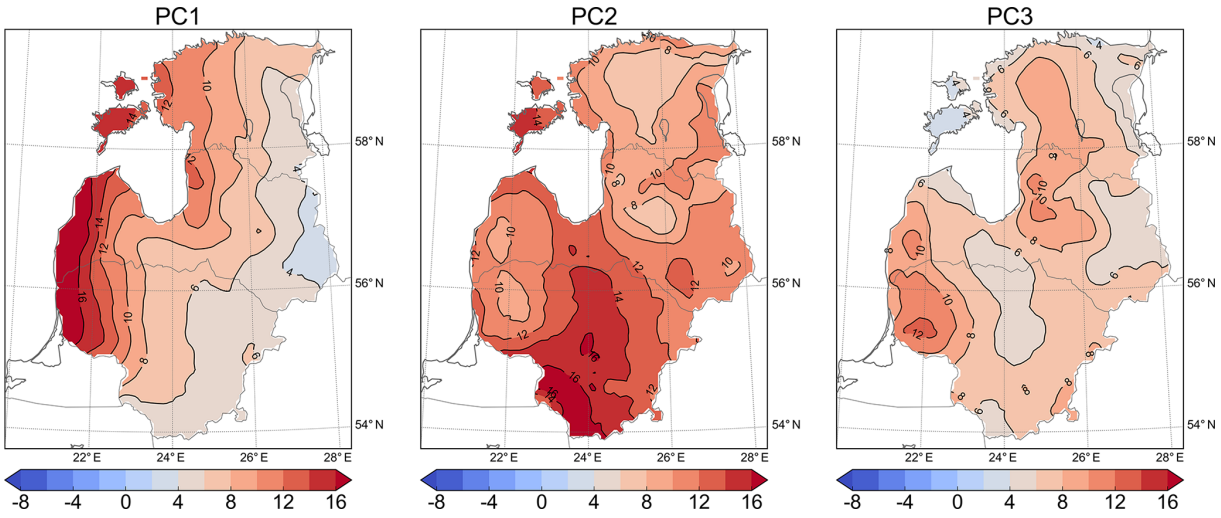

Figure 7. Climate indices (based on principal components from 1961–1990) for the years 2071–2100.

Table 6. Statistics of climate indices (based on PCA) for past and future data.

		1961–1990	2071–2100
PC1	Mean	0.00	8.38
	Min	−4.84	3.17
	Max	8.95	18.24
PC2	Mean	0.00	11.38
	Min	−5.62	6.24
	Max	6.14	17.05
PC3	Mean	0.00	7.13
	Min	−8.43	1.54
	Max	4.84	12.28

All indices have higher values in future climate. This can be interpreted as an overall warmer climate (increase in PC2) and wetter climate (increase in PC3). The interpretation of PC1 is more complicated as coefficients (Table 4) for some variables are positive and negative for others. An increase in PC1 would be observed in the case of a constant increase in all variables. However, an increase would also be observed in the case of a temperature and precipitation decrease in spring and summer. An average increase of "standardized" (by Eq. 5) mean values is 1.4 units for temperature and 4.5 units for precipitation. Such a constant increase with the coefficients in Table 4 would result in a 6.5 unit increase for PC1. As we can see from the index statistics in Table 7, an increase of 8.4 units is observed for PC1, so we suspect that the additional increase can be attributed to changes in seasonality.

For PC1 it is shown that the values corresponding to coastal regions in the reference climate will "move" to the eastern part of the Baltic states in the future projections. The expected changes in PC2 are the largest, and the maximum values of PC2 for the reference climate (in southern Lithuania) are lower than the minimum values for the future climate (in central Estonia). The statistics in Table 6 show that the reference range of this index does not overlap with the range of future values. The climate corresponding to the reference values of PC3 in western Lithuania (the Zemaiciai Highland) will in the future be observable on plateaus in the central and northeastern parts of the Baltic states.

4 Discussion

The methodology used in this study has been able to reduce 24 climate variables to three new indices that more efficiently and compactly represent the main features of the climate in the Baltic countries. The methodology can also be applied to future climate data and therefore the impacts of climate change can be analyzed. Additional analysis is needed for the interpretation of the acquired indices.

Some insight into the possible interpretation of the acquired climate indices can be gained from the literature. The spatial distribution of PC1 is similar to the spatial patterns of the mean start date of winter (see results for Estonia in Jaagus and Ahas, 2000) with higher PC1 values corresponding to later winters.

As PC2 is mainly linked to temperature, the patterns exhibited by PC2 can be expected to be similar to the spatial distribution of phenological events for which temperature is the main driving factor. For example, the spatial pattern of PC2 shows similarities to spring and summer start dates in the Baltic Sea region and to more specific phenological events, such as apple tree blossoming and the beginning of the vegetation of rye (Jaagus and Ahas, 2000) or strawberry blooming and harvest (Bethere et al., 2016). In gen-

Table 7. Description and interpretation of climate indices based on PCA.

Name	High values correspond to locations with	Possible interpretation of high values
PC1	Warm winter with high precipitation, cold summer with low precipitation	Less distinct seasonality
PC2	High overall temperature, low precipitation in autumn	Warmer climate
PC3	High annual precipitation, warmer springs	More humid climate

eral, higher values of PC2 correspond to places with earlier phenological processes.

High values of winter precipitation and high temperatures in spring can be interpreted in the context of spring floods; however, additional analysis is needed to account for the snow cover. The spatial distribution of PC3 is similar to the map of average annual precipitation in the study region (Jaagus et al., 2010). Interestingly, the precipitation in autumn months (September–October) has a small contribution to PC3 (Table 5).

Conclusions based on spatial pattern and correlation coefficient analysis are summarized in Table 7.

The methodology could be further improved to better link the acquired indices with phenological processes or seasons by either rotating the acquired principal components (Jolliffe, 2002) or performing correlation or regression analysis with other variables, such as crop yield (Cai et al., 2013). This approach would be especially useful in the case of PC1, for which analysis is currently complicated due to both changes in seasonality and the constant increase affecting PC1 values. Another approach that could be used to describe the spatial variability of the climate in the Baltic states is clustering based on the chosen principal component values (Fovell and Fovell, 1993; Forsythe et al., 2015).

If variables other than temperature or precipitation are used for the principal component analysis, in some cases the standardization procedure should be modified. However, it should be taken into account that when more than one data set is used, e.g., when past and future climate is compared, the same values used for standardization should be applied to all of them.

5 Conclusions

Most of the spatial variability in monthly average temperature and precipitation over the Baltic countries can be represented by three principal components for both past and future climate. These components can be considered climate indices, in which higher values correspond to locations with (1) climate with less distinct seasons, (2) warmer climate, and (3) climate with more precipitation. Each component has a distinct spatial pattern. The index related to seasonality exhibits a clear east–west (or inland) gradient with less distinct seasonality at the seaside (west). The second index (warmer climate) shows a north–south gradient with a warmer climate in the south. This index also reflects orography with colder climate in hilly regions. The third index reflects the overall precipitation. Its spatial distribution is mainly dominated by elevation, with maxima at the highlands and less precipitation in the plains and at the seaside. A specific standardization of the data also allows for the calculation of such indices for the future climate. Change in the climate indices in the future implies less distinct seasons and a warmer and wetter climate.

Although there is significant change in the magnitude of the indices between the future and reference periods, the change in spatial distribution is relatively small. For the first and third components, regions can be identified in which the future climate will be similar to the current climate in other regions.

Data availability. We used publicly accessible data (before bias correction; the bias correction is described in detail in Sennikovs and Bethers, 2009). RCMs are from the ENSEMBLES project: http://ensemblesrt3.dmi.dk/. Observations (combination of two sources) from the Latvian Environment, Geology and Meteorology Centre: https://www.meteo.lv/en/meteorologija-datu-meklesana/?nid=924. Europe Climate Assessment & Dataset project: http://www.ecad.eu/dailydata/index.php.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue "Multiple drivers for Earth system changes in the Baltic Sea region". It is a result of the 1st Baltic Earth Conference, Nida, Lithuania, 13–17 June 2016.

Acknowledgements. The research was supported by the Latvian state research program "The value and dynamic of Latvia's ecosystems under changing climate" (EVIDEnT).

The ENSEMBLES data used in this work were funded by the EU FP6 Integrated Project ENSEMBLES (contract number 505539), and support is gratefully acknowledged.

Edited by: Anna Rutgersson
Reviewed by: two anonymous referees

References

BACC II: Second assessment of climate change for the Baltic Sea basin, Reg. Clim. St., Springer, https://doi.org/10.1007/978-3-319-16006-1, 2015.

Baeriswyl, P. A. and Rebetez, M.: Regionalization of precipitation in Switzerland by means of principal component analysis, Theor. Appl. Climatol., 58, 31–41, https://doi.org/10.1007/bf00867430, 1997.

Belda, M., Holtanová, E., Halenka, T., and Kalvová, J.: Climate classification revisited: from Köppen to Trewartha, Clim. Res., 59, 1–13, https://doi.org/10.3354/cr01204, 2014.

Benzi, R., Deidda, R., and Marrocu, M.: Characterization of temperature and precipitation fields over Sardinia with principal component analysis and singular spectrum analysis, Int. J. Climatol., 17, 1231–1262, https://doi.org/10.1002/(sici)1097-0088(199709)17:11<1231::aid-joc170>3.3.co;2-1, 1997.

Bethere, L., Sile, T., Sennikovs, J., and Bethers, U.: Impact of climate change on the timing of strawberry phenological processes in the Baltic States, Est. J. Earth Sci., 65, 48–58, https://doi.org/10.3176/earth.2016.04, 2016.

Bieniek, P. A., Bhatt, U. S., Thoman, R. L., Angeloff, H., Partain, J., Papineau, J., Fritsch, F., Holloway, E., Walsh, J. E., Daly, C., and Shulski, M.: Climate divisions for Alaska based on objective methods, J. Appl. Meteorol. Clim., 51, 1276–1289, https://doi.org/10.1175/jamc-d-11-0168.1, 2012.

Briggs, R. D. and Lemin Jr., R. C.: Delineation of climatic regions in Maine, Can. J. Forest Res., 22, 801–811, https://doi.org/10.1139/x92-109, 1992.

Cai, R., Mullen, J. D., Bergstrom, J. C., Shurley, W. D., and Wetzstein, M. E.: Using a climate index to measure crop yield response, J. Agr. Appl. Econ., 45, 719–737, https://doi.org/10.1017/s1074070800005228, 2013.

Cattell, R. B.: Factor analysis: an introduction and manual for the psychologist and social scientist, Harper, New York, 1952.

De Castro, M., Gallardo, C., Jylha, K., and Tuomenvirta, H.: The use of a climate-type classification for assessing climate change effects in Europe from an ensemble of nine regional climate models, Climatic Change, 81, 329–341, https://doi.org/10.1007/s10584-006-9224-1, 2007.

Estrada, F., Martinez-Arroyo, A., Fernández-Eguiarte, A., Luyando, E., and Gay, C.: Defining climate zones in Mexico City using multivariate analysis, Atmosfera, 22, 175–193, 2009.

Fan, F., Bradley, R. S., and Rawlins, M. A.: Climate change in the northeastern US: regional climate model validation and climate change projections, Clim. Dynam., 43, 145–161, https://doi.org/10.1007/s00382-014-2198-1, 2014.

Feng, S., Ho, C. H., Hu, Q., Oglesby, R. J., Jeong, S. J., and Kim, B. M.: Evaluating observed and projected future climate changes for the Arctic using the Köppen-Trewartha climate classification, Clim. Dynam., 38, 1359–1373, https://doi.org/10.1007/s00382-011-1020-6, 2012.

Forsythe, N., Blenkinsop, S., and Fowler, H. J.: Exploring objective climate classification for the Himalayan arc and adjacent regions using gridded data sources, Earth Syst. Dynam., 6, 311–326, https://doi.org/10.5194/esd-6-311-2015, 2015.

Fovell, R. G. and Fovell, M. Y. C.: Climate zones of the conterminous United States defined using cluster analysis, J. Climate, 6, 2103–2135, https://doi.org/10.1175/1520-0442(1993)006<2103:CZOTCU>2.0.CO;2, 1993.

Hagen, E. and Feistel, R.: Climatic turning points and regime shifts in the Baltic Sea region: the Baltic winter index (WIBIX) 1659–2002, Boreal Environ. Res., 10, 211–224, https://doi.org/10.1109/baltic.2014.6887870, 2005.

Jaagus, J. and Ahas, R.: Space–time variations of climatic seasons and their correlation with the phenological development of nature in Estonia, Clim. Res., 15, 207–219, https://doi.org/10.3354/cr015207, 2000.

Jaagus, J., Briede, A., Rimkus, E., and Remm, K.: Precipitation pattern in the Baltic countries under the influence of large-scale atmospheric circulation and local landscape factors, Int. J. Climatol., 30, 705–720, https://doi.org/10.1002/joc.1929, 2010.

Jolliffe, I.: Principal component analysis, Springer, New York, 2002.

Karl, T. R., Koscielny, A. J., and Diaz, H. F.: Potential errors in the application of principal component (eigenvector) analysis to geophysical data, J. Appl. Meteorol., 21, 1183–1186, https://doi.org/10.1175/1520-0450(1982)021<1183:peitao>2.0.co;2, 1982.

Mahlstein, I. and Knutti, R.: Regional climate change patterns identified by cluster analysis, Clim. Dynam., 35, 587–600, https://doi.org/10.1007/s00382-009-0654-0, 2010.

Mahlstein, I., Daniel, J. S., and Solomon, S.: Pace of shifts in climate regions increases with global temperature, Nat. Clim. Change, 3, 739–743, https://doi.org/10.1038/nclimate1876, 2013.

Malmgren, B. A. and Winter, A.: Climate zonation in Puerto Rico based on principal components analysis and an artificial neural network, J. Climate, 12, 977–985, https://doi.org/10.1175/1520-0442, 1999.

Netzel, P. and Stepinski, T.: On using a clustering approach for global climate classification, J. Climate, 29, 3387–3401, https://doi.org/10.1175/jcli-d-15-0640.1, 2016.

Overland, J. E. and Preisendorfer, R. W.: A significance test for principal components applied to a cyclone climatology, Mon. Weather Rev., 110, 1–4, https://doi.org/10.1175/1520-0493(1982)110<0001:astfpc>2.0.co;2, 1982.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen–Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, https://doi.org/10.5194/hess-11-1633-2007, 2007.

Preisendorfer, R. W. and Mobley, C. D.: Principal component analysis in meteorology and oceanography, Elsevier, Amsterdam, 1988.

Remm, K., Jaagus, J., Briede, A., Rimkus, E., and Kelviste, T.: Interpolative mapping of mean precipitation in the Baltic countries by using landscape characteristics, Est. J. Earth Sci., 60, 172–190, https://doi.org/10.3176/earth.2011.3.05, 2011.

Rutgersson, A., Jaagus, J., Schenk, F., and Stendel, M.: Observed changes and variability of atmospheric parameters in the Baltic Sea region during the last 200 years, Clim. Res., 61, 177–190, https://doi.org/10.3354/cr01244, 2014.

Sennikovs, J. and Bethers, U.: Statistical downscaling method of regional climate model results for hydrological modelling, Proc. 18th World IMACS/MODSIM Congress, Cairns, Australia, 13–17, 2009.

Tapiador, F. J., Angelis, C. F., Viltard, N., Cuartero, F., and De Castro, M.: On the suitability of regional climate models for reconstructing climatologies, Atmos. Res., 101, 739–751, https://doi.org/10.1016/j.atmosres.2011.05.001, 2011.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods, J. Hydrol., 456, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

Van der Linden, P. and Mitchell, J. E.: ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, Exeter, 160, 2009.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Vol. 100, Academic Press, Elsevier, SanDiego, CA, USA, 2011.

Zhang, X. and Yan, X.: Spatiotemporal change in geographical distribution of global climate types in the context of climate warming, Clim. Dynam., 43, 595–605, https://doi.org/10.1007/s00382-013-2019-y, 2014.