Thank you for the reviews. When writing the revised version we carefully considered each of the comments made by referees. This file contains point-by-point response of initial referee reviews and in **orange** are written additional comments explaining how we addressed the comment in revised version of the paper.

Referee #1

1. The abstract is too general. I would like to see more concrete results of the study in the abstract.

Will be considered in revision.

We added additional details to abstract both clarifying methodology and expanding on results.

2. Page 1 line 29. What does mean here the term "homogeneous"? How can the results be more homogeneous? In which sense?

In the mentioned article (Netzel and Stepinski, 2016) it was obtained that "we demonstrate that clustering-based classification results in climate types that are internally more homogeneous and externally more distinct than climate types in the KGC". Homogenity was measured as "For each class the entropy of a histogram of it constituent types measures a level of class homogeneity with respect to types. Homogeneous classes (like A) have small values of entropy and inhomogeneous classes (like D) have large values of entropy"

We don't think that introduction of the article should include definition of homogenity. So no changes were done regarding this comment.

3. Page 2 lines 3-5. This sentence is a bit unclear and confusing to me. It is written the loadings of components are the coefficients that define indices. I have an imagination that loadings of principal components show correlation between time series of the components and observed variables. Can you explain this?

"The loadings of chosen principal components are the coefficients that define the newly created indices, which then describe the main features of climate."

We have chosen to use most common terminology (at least this is mentioned as most common terminology in Wilks, 2006. Statistical methods in the atmospheric sciences. This book is not mentioned in references, as this terminology is also used in book by Jolliffe, 2002, that we use).

Values of principal components are calculated as P = Xe, where X is data matrix of initial data, e are eigenvectors or loadings and P are principal components. Explained variance is calculated from eigenvalues.

This means, that correlation coefficients are: corr.coef = loadings * sqrt(eigenvalues).

Also we will specify used terminology in revision.

We have explained methodology in more detail and clarified the terminologu that we use.

4. Page 2 line 24. Should it be the reference to de Castro et al. 2007?

"RCM models are continuously improving and correspond rather well to climate observations (Castro et al. 2007)."

We will revise to make it accurate.

There indeed was mistake in references, this part of the article has been revised and corrected.

5. The introduction is lacking of the description of similar studies. PCA is widely used in climatology and also for determining of various climate indices. I suggest a literature overview where is shown how the current study is fitting into other similar studies. The novelty of this study should be clearly indicated.

Will be considered in revision.

We feel that we have covered most notable articles in regard to each of the aspects of our study in initial version of the paper already:

- PCA application in general;

- PCA methodology;

- PCA use in index development;

- PCA in Baltic countries;

- use of RCM data.

Our publication is new in combining each of these aspects in a way that hasn't been done before. No changes were made to the paper regarding this comment.

6. The description of the use of model data could be more precise. Does the ensemble data mean that the averages of 22 model runs were calculated? What was the spatial resolution of the ensemble data? I suppose that the resolution was different for every single model run.

We will revise to make it accurate.

The processing of ensemble data is fully based on publication by Sennikovs and Bethers, 2009 and process is described in more detail there. When revising the paper we decided to emphasize more the reference in our paper instead of fully copying parts of referenced paper.

7. It was not clear why the model data were used instead of station data. The density of meteorological stations is rather high in the study area. Therefore, the results of the PCA of station data would be compared with the results of the RCM-based data.

Main reason was because we have model data for future period. Also model data was bias corrected based on station data so statistics for each case (and also PCA result) should coincide.

No changes were made in the paper regarding this comment.

8. Page 3 line 9. It is not indicated from which source the observation data (Fig.1) were obtained.

Bias-correction was fully described in (Sennikovs and Bethers 2009). We are considering to remove the illustration of locations and put more emphasis on reference.

See answer to comment 6.

9. Page 3 line 18. Usually, it is written "…as it is done by Malmgren et al. (1999) and Forsythe et al. (2015)".

Will be considered in revision.

Was corrected as suggested.

10. Page 3 lines 22-23. A very strong correlation in winter precipitation is detected. Is it so that correlations are calculated using the data from the same year? In that case there is not any time lag. I don't believe that there is a correlation above 0.8 between January and December of the same year. I don't believe the statement that winters are either dry or humid. There should be something wrong. I did some calculations with station data of monthly precipitation and did not find any significant correlations. Correlations presented in Fig. 4 are inadequate. Such high correlations in monthly precipitation are not possible at all. All other correlation coefficients seem also suspicious. The reason for presenting the correlation Matrix is not clear.

We use climatic variables – 30 year average. Data matrix that is used for both correlation calculation and PCA consists of 24 variables (12 temperature and 12 precipitation values) and 7143 cases (grid points). It means that we are looking for spatial patterns, not temporal. This approach is less used, but there are similar applications in literature (for examples, Fovell and Fovell 1993).

The reason for correlation matrix was to show that there is redundant information that should be reduced through PCA.

There was no mistake in our calculations and we have clarified the methodology in the paper.

11. Page 4 line 18. I suggest the word "them" instead of "then".

We will revise to make it accurate.

Was corrected as suggested.

12. Standardisation of climatic data is a trivial procedure. It is not clear why the variances on tables 2 and 3 are presented in the study. Were they spatial mean variances?

Standardization of data is an important part of PCA that can influence acquired result, and as we are using a bit different approach from the standard (standard meaning subtracting the mean of variable and dividing by square root of variance) it is important to both clarify what we're doing and why. Main reason for tables 2 and 3 was to show the variances for

each variable and the mean variance per variable category (temperature or precipitation) after we had performed or standardization procedure. Table 2 illustrates what we tried to accomplish (and succeeded) in our standardization procedure. Table 3 illustrates differences in future data (from reference period data in Table 2) in regard to variance of data. Table 3 shows increased precipitation data variance and reduced temperature variance in comparison to reference period. Also Table 3 shows that pattern of variances between different months haven't changed (I was considering illustrating this point, but this would just duplicate information in Tables 2 and 3).

I hope that our explanation has clarified everything. In paper changes were minor, we slightly emphasized the impact of standardization on PCA.

13. I think that more information about the PCA procedure is needed. Was it rotated or non-rotated PCA? There are different modes of PCA: T-mode, S-mode etc. How the matrix was performed? Which were variables and which were cases? What were loadings and what were scores? This information is needed for the interpretation of results.

Will be elaborated in revision. We used non-rotated PCA, and matrix is described in point 10 of this document. Terminology we used is shortly summarized in point 3 of this document. (Principal components = scores. Loadings = eigenvectors). Also we will consider specifying used terminology and technique in revision.

Additional description of methodology was added to the paper.

14. Why the loadings of the three first components are presented in Table 4. What do they show?

They define approach of using PCA results as climate indices. (This is elaborated a bit more in point 19 of this document).

We have described methodology and terminology of index calculation in more detail.

15. The spatial patterns of three first components are very interesting and informative. I think that they could be wider and better interpreted. It is clear that PC1 represents the influence of the Baltic Sea. It is the main factor causing spatial climatic differences in the Baltic countries. It is directly related to higher temperature and precipitation in autumn and winter, and lower temperature and precipitation in spring and early summer in the coastal regions. In the hinterland far from the sea the spatial coefficients (scores ?) are negative. In conclusion, PC1 reflects continentality of climate.

Useful input, will be taken into account in revision. It is important to note that values itself (negative, positive, etc.) don't have a meaning associated with them! For example our standardization process of subtracting mean resulted in negative values for precipitation. If we used a different approach, for example, subtracting minimum value, we would get different values of principal components (scores). However that wouldn't change interpretation (correlation coefficients) of PCA results or spatial pattern. What we can compare are regions (or their lack) with similar values.

Some of the points made by referee regarding the interpretation were included in the revised version of the paper.

16. PC2 reflects the second main factor in formation of climate – i.e. latitude. The pattern shows positive scores in Lithuania and negative scores in Estonia. The southern region is characterised by higher temperature, especially in spring and autumn, comparatively higher precipitation from April to June and lower precipitation during the rest of a year.

Useful input, will be taken into account in rewrite. About positive/negative values see point 15 of this document.

Some of the points made by referee regarding the interpretation were included in the revised version of the paper.

17. The spatial pattern of the PC3 is very similar to the mean annual distribution of precipitation in the study region (Jaagus et al. 2010). Two regions with higher precipitation are described by areas of negative coefficients - one in western Lithuania and Latvia, and another in the western part of continental Estonia and central Vidzeme upland in Latvia. Positive areas correspond to coastal regions with lower precipitation in Estonia and Latvia. But I cannot understand why spatial coefficients (loadings) on Fig. 6 are negative but temporal coefficients (scores) in Table 5 are positive. I cannot fully understand the results of PCA.

Useful input, will be taken into account in rewrite. Due to standardization some values of precipitation are negative, that can result in negative coefficients. About positive/negative values see point 15 of this document.

Some of the points made by referee regarding the interpretation were included in the revised version of the paper.

18. I suggest that the authors do not interpret the results fully and not always adequately. If I understand correctly, interannual variations of temperature and precipitation are not reflected in the results of PCA. There are presented only mean monthly variability. Consequently, the results of current PCA reflect spatio-temporal variability of monthly mean values. Therefore, the interpretation of the results on page 5 is not valid in the following sentences: "This means that higher values of PC1 correspond to warmer winters …" (lines 15-16), "In general, higher values of PC2 correspond to earlier phenological processes" (lines 28-29), "This means that high PC3 values correspond to overall high precipitation and warm spring, or in other words – overall wetter year" (lines 31-32). If interannual variability is not included into the analysis, the relationships with phenological phases are not appropriate. Anyway, here are many problems to be clarified.

We will rephrase our interpretations in revision to avoid confusion.

We have carefully reviewed our interpretation of the results and clarified it wherever possible – mostly emphasizing that correlation coefficients and PCA results describe differences between locations (similarities/differences are spatial not temporal).

19. It is not clear how the loadings were used to calculate climate indices for the future. I am not sure but it would be correct to realise PCA for the modelled mean values for 2071-2100 and analyse the results of past and future analyses.

One of the aims of this work was to see how our components (climate indices) change in future.

I will try to better explain reasoning for method used in this paper. So the idea is that we perform PCA and acquire principal components, and then once we have some kind of interpretation of principal component, we can just assume that it's a climate index. Try to discard for a moment the principal componets part and just think about the climate index

$$CI = a_1T_1 + \cdots + a_{12}T_{12} + a_{13}P_1 + \cdots + a_{24}P_{12}$$

We have these coefficients a_1 … a_{24} and we can just calculate this index for present and future.

An analogy would be for example growing degree days. If we want to compare change of growing degree days in present and future we would want to use same base temperature and calendar period to calculate the sum of growing degree days. Even though there might be an argument that in future there is shift in seasons or plant adaptation that would affect methodology. Similarly in our case, we wish to use the same coefficients a_1 … a_{24} for reference period and future, because only then can we make conclusions about the change in value of chosen climate index (It's important to note that currently we can make conslusions about general increase of climate index value or identify regions that have similar values. However we can't make any conclusions about numerical increase, for example, what is the meaning of increase of 5 or 10 units) . Because of the issue about comparison also there were some considerations about standardization process (especially application to future).

We have described methodology and terminology of index calculation in more detail.

20. The section of discussion is pure. I recommend to restructure the paper. In the section of results, there could be only the description of the results of PCA. All interpretations might be included into the section of discussion.

We acknowledge the problem and will consider in it revision.

We have revised and restucturized 'Results' and 'Discussion' part of the paper.

Referee #2

Responses to the comments:

1. **It was mentioned by the referee that scaling should be reconsidered.**

Removal of seasonal cycle doesn't impact our result as PCA is performed on covariance matrix. For example, let's look at two variables X and Y. Let's assume we are performing some standardization on each of them and acquiring new variables X' and Y':

$$X' = \frac{X - C_1}{S_1}, \;\; Y' = \frac{Y - C_2}{S_2}$$

Covariance between initial variables is:

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y})}{n-1}$$

Covariance between transformed variables:

$$Cov(X',Y') = \frac{\sum_{i=1}^{n}(X_i{}' - \bar{x}')(Y_i{}' - \bar{y}')}{n-1}$$

Where:

$$\bar{x}' = \frac{1}{n}\sum_{i=1}^{n} X_i' = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - C_1}{S_1}\right) = \frac{1}{S_1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \; C_1\right) = \frac{1}{S_1}(\bar{x} - \; C_1)$$

And:

$$(X_i' - \bar{x}') = \; \frac{1}{S_1}(X_i - C_1 - \bar{x} + \; C_1) = \frac{1}{S_1}(X_i - \bar{x})$$

Similarly:

$$(Y_i' - \bar{y}') = \frac{1}{S_2}(Y_i - \bar{y})$$

And this implies that covariance and therefore PCA is not affected by subtracted values:

$$Cov(X',Y') = \frac{\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y})}{(n-1)S_1 S_2}$$

Subtraction of mean values is important for visualization as it gives similar range for principal components and therefore of their illustration. The change of scaling in regard to subtraction of mean value will impact the values, but won't impact the pattern.

As for scaling future – we wish to compare the change in climate index from past to future and data processing should be kept similar in our case. As a similar example could be considered comparison of growing degree day sum in past and future. If we wish to make comparison then usually same base temperature and period (start, end date) should be used for both past and future, even though there might be changes in seasonality.

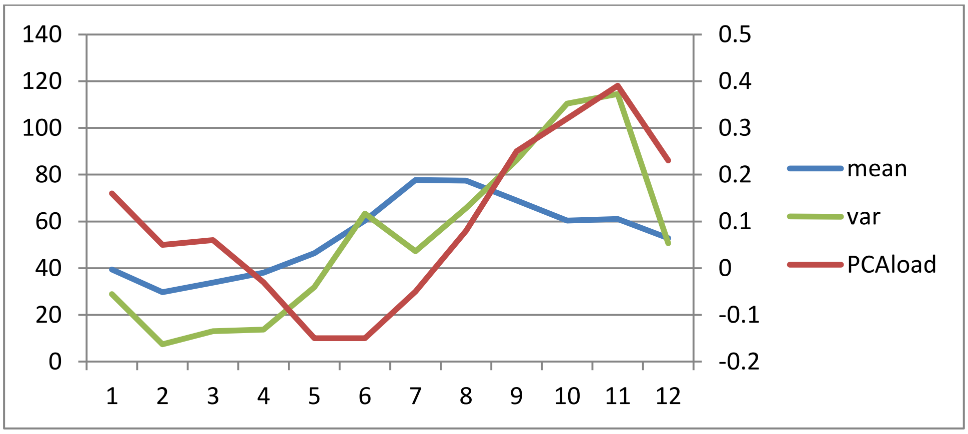
Also our method was mentioned as valid in Cai et al. (2013). Still we agree that adjustments to scaling should be considered (in addition to rotation of principal components) in future development of these climate indices and this point was mentioned in the 'Discussion' part.

We have improved methodology description and description of motivation for used scaling methodology to avoid confusion.

2. Result interpretation and conclusions

We agree that the interpretation of our results should be reviewed and often clarified or reconsidered, especially that it should be emphasized that any correlation we mention is spatial not temporal. Many of referee's comments are on point and will be implemented in revised version of the paper.

"The difference is clearer for precipitation, where the PC1 precipitation loadings have a max in November and a min in May/June, whereas the precipitation seasonal cycle has a max in August and a min in February. That is, the annual cycle of the precipitation distance-to-the-coast effect is nearly orthogonal to the season cycle of precipitation itself – high values of PC1 do not describe a climate in which precipitation is similar throughout the year!"



The aim of PCA is to explain variance, so there is high correlation with seasonal variance, not seasonal mean values.

We have reviewed description and interpretation of correlation calculation and PCA results and clarified that we are talking about spatial not temporal correlation and differences/similarities. We think that with these changes the aim and results of the paper agree with each other and no significant changes are required for either aim or methodology.

In addition referee mentioned the orthogonality between precipitation seasonal cycle and PC1. However as we have shown in previous response (first point), seasonal cycle doesn't

impact PC1 loadings, so independence can be expected. Instead loadings are based on the variances (as reflected in the graph above).

3. **Details on interpolation methodology. "Finally, the methods section is missing important information: What interpolation method was used? What grid did you interpolate onto? what resolution ENSEMBLES simulations were used? did you bias-correct each pixel from each RCM separately? How did you deal with cases where there are more than one station corresponding to a given RCM pixel (if you used 50km resolution simulations, this must have happened a lot?)"**

This is explained in detail in the article that we referenced: Sennikovs, J., Bethers, U.: Statistical downscaling method of regional climate model results for hydrological modelling, Proc.18th World IMACS / MODSIM Congress, Cairns, Australia 13-17, 2009.

We will consider adding more information about the methodology in the revised version of the article.

The processing of ensemble data is fully based on publication by Sennikovs and Bethers, 2009 and process is described in more detail there. When revising the paper we decided to emphasize more the reference in our paper instead of fully copying parts of referenced paper.

4. **"Table 4, 5 and Figure 7 contain essentially the same information; not sure it is worth having all three."**

Will be considered in the revised version. There is purpose for each table/figure. Table 4 defines climate indices (this table probably can be considered for removal). In table 5 correlation coefficients are calculated. They are similar to loadings, but there are differences that make them more suited for interpretation of the results. And figure 7 is required to show that acquired climate indices hold there meaning also in future (alternative was table of correlation coefficients for future, but figure was both more illustrative and felt less redundant). We will consider adding more explanation in revision of paper on use of correlation coefficients.

We have added more explanation of methodology and differences between loadings, correlation coefficients etc. This should clarify usefulness of each table.

Major changes:

1. More comprehensive description of methodology is added to clarify terminology and avoid confusion.
2. Method of calculating climate indices has been expanded and clarified.
3. Restructuration of results and discussion parts of the article has been done.
4. Clarification of results and their interpretation (main purpose: emphasize that spatial pattern is detected and analyzed) has been done.
5. We noticed a mistake in Figure 5. (previously Figure 6.) and the figure was replaced. There is no change in any conclusions linked to that figure;
6. Referees raised several questions in regard to methodology of RCM data use. The method is fully based on the article by Sennikovs and Bethers (2009) and is referenced in the paper. Figure 1 (bias correction locations) is removed as article by Sennikovs and Bethers contains similar, but more informative figure.

Climate indices for Baltic States from principal component analysis

Liga Bethere[1], Juris Sennikovs[1], Uldis Bethers[1]

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

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

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 (RCM) were used. Principal components were derived for years 1961-1990. The first three components describe 92% of the variance of 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 ~~observed~~detected (based on correlation coefficient between principal components and initial variables) that higher values of 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 impact of the Baltic Sea (distance to coast) was apparent, for the second – latitude and elevation, and for the third - elevation. The loadings from the chosen principal components were ~~then~~ further used to calculate values of the climate indices for years 2071-2100. Overall increase was found for all three indices with minimal changes in their spatial pattern.

1 Introduction

Spatial representation of the climate e.g. 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 both into the possible relationship and the impacts of the climate to other fields, e.g., phenological processes and vegetation distribution (Feng et al. 2012). Third, they illustrate geographical features that influence climate, such as hillsides, coastal zones etc. There is a wide variety of approaches for creating spatial representation 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 Köppen-Geiger classification (Peel et al. 2007). These methods are based on some 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. 2008) and 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 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~~PCA~~ provide~~s~~ new indices that are the linear combination of chosen variables. The loadings of 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 climate of the region in general (Estrada et al. 2008). 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 cycling of nutrients are strongly influenced by temperature and moisture (Briggs and Lemin 1992). Distinct changes of temperature and precipitation are to be expected in future (BACC 2015). Thus, consequently, any climate patterns based on these two variables also will be affected, leaving 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 States region exhibits significant spatial and temporal climatic variability, with influence of air masses from 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 meters. 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 pattern in the Baltic countries with atmospheric and landscape variables (Jaagus et al. 2010).

To study 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). RCM models are continuously improving and correspond rather well to climate observations (Tapiador et al. 2011~~Castro et al. 2007~~). Other advantages of using RCM data are that (a) their data are regularly spaced while PCA applied on irregularly spaced data can produce distorted loading patterns (Karl et al. 1982) and (b) RCM data are available also as future projections giving insight into 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 which represent the main features of Baltic States 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 analysis was performed and main principal components were identified. Acquired principal components and their spatial patterns were analyzed. Fourth, loadings of chosen principal components were used to calculate indices for years 2071-2100 and compared to reference data.

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. 22 model runs were considered (shown in Table 1).

We used time series of daily average air temperature at 2m height and daily precipitation. RCM models are known to show systematic biases (Teutschbein & Seibert 2012). 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 11-day running average – the data for five days before and five days after the day of interest. The control period for bias correction was 1961-1990~~ and the locations of observation stations used for bias correction are shown in Fig. 1~~. Bias corrected data was 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). Bias correction method and model resolution is described in detail in Sennikovs and Bethers, 2009.

Two time periods were chosen – 1961-1990 (as 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 reference period is shown in Fig. ~~2~~ 1 and Fig. ~~3~~2. In Fig.~~ure~~ ~~2~~ 1 ~~shows~~can be seen north-south gradient of monthly precipitation during April-June and east-west gradient of monthly precipitation during October-January. Figure ~~3~~ 2 shows east-west gradient of monthly temperatures during October-February and north-south gradient of monthly temperatures during April-June. This implies that some of the variables can be combined in seasons (as it is done by ~~(~~Malmgren et al. (1999)~~)~~ and ~~(~~Forsythe et al. (2015)~~)~~) and that for some months temperature and precipitation is correlated. A better insight of variables with similar patterns can be gained by examining the correlation matrix in Fig. ~~4~~3. The matrix areas that represent strongly correlated variables are marked in this figure and they show following relationships:

1 - Very strong correlation (above 0.8) between precipitations in winter months – locations with more precipitation in, e.g., December also have ~~is clearly linked to~~ more precipitation in January (compared to the rest of the territory).~~ Thus, entire winters are either dry or humid.~~

2 - Strong correlation (above 0.5) between precipitation and temperature in spring months. Thus, locations with colder springs ~~are~~ also are dryer, whilst locations with warmer springs also have more spring precipitation~~ are rainy~~.

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~~corresponds to~~ colder spring.

4- Very strong correlation (above 0.8) between temperatures of autumn and winter months – locations with warmer autumn also have ~~corresponds to~~ warmer winter.

Figure ~~4~~ 3 shows that 24 monthly variables contain redundant information and through PCA we can summarize information and create new variables.

2.2 PCA method

The aim of PCA is to create a new set of uncorrelated variables that are linear combination of initial variables and explain as much as possible of the initial variation. 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 literature can vary (Wilks, 2011). We will briefly describe the terminology used in this article.

Suppose that 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). Then by solving Eq. (2) we can find eigenvectors $e_i, i = 1, \dots, 24$ and corresponding eigenvalues $\lambda_i,\ i = 1, \dots, 24$. As a result we have obtained non-correlated linear combinations of initial climatic variables calculated by Eq. (3).

$$S = (n-1)^{-1} X^T X. \qquad (1)$$

$$Se = \lambda e \qquad (2)$$

$$Y_i = X_i e_i,\ \ i = 1, \dots, 24. \qquad (3)$$

Values λ_i represent explained variance of each “principal component” Y_i. Linear weights e_i that define each principal component will be called “loadings”. “Indices” describe Y_i that are calculated using loadings from 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 correlation matrix or covariance matrix in the calculation of loadings~~coefficients that define principal components (eigenvectors)~~. If the covariance matrix is used then a second choice must be made – if and what standardization to use. Scaling process has a significant impact on the PCA process. When performing data standardization following issues should be taken into account:

1 – Variables should be of 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 of temperature that is measured in degrees Celsius.

2 – In case of variables that are measured in same units variances contain useful information and can improve interpretation of PCA (Overland and Preisendorfer 1982). Therefore, for variables that are measured in same units (for example, average temperature of different months) we wish to keep the ratio between variances of different months. This means that

correlation matrix, where 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 ~~for climate index calculation for future data~~ and compare the~~n~~m with the reference climate it is necessary that the same standardization process is used for the data of the future time period.

4 – It is important to note that subtraction of mean (or 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 zero mean (mean is subtracted from each variable) then resulting principal components have similar scale and spatial patterns are more convenient to review.

Taking into account the issues described above we propose to use standardization ~~in~~ as defined by Eq. (~~1~~4), where the spatial mean is subtracted for each variable as usual, but the average variance of all temperature or precipitation variables is used for scaling:

$$\frac{T_k-\bar{T}_k}{\sqrt{\bar{V}(T)}}, \qquad \frac{P_k-\bar{P}_k}{\sqrt{\bar{V}(P)}}, \qquad k = 1, \dots ,12, , \qquad (\cancel{1}4)$$

w~~W~~here $\bar{V}(T), \bar{V}(P)$ – average variance of 12 temperature and precipitation variables for reference period~~$\bar{V}(T)$ – average variance of 12 temperature variables and $\bar{V}(P)$ – average variance of 12 precipitation variables~~.

The variances before and after such standardization for 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 period. Results of data standardization for future time period are shown in Table 3. It can be seen ~~should be noted~~ that in the future the variance of precipitation data will increase and the variance of temperature data will decrease. However, the distribution of variances over the year ~~proportion of variances in different months~~ 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 – scree-plot – gives excellent and clear results. A scree plot is a graph of explained variances of 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)

Explained variance and loadings of first 3 principal components are shown in Table 4. The scree-plot of all principal components is shown in Fig. ~~5~~4. First two components already describe 78% of variance of initial variables, while first three components describe 92% of variance. According to Jolliffe (2002) the cutoff point should be between 70% and 90% of explained variance. However, the scree-plot clearly shows that first 3 principal components can be retained, so we chose to further analyze first 3 components.

Figure ~~6~~ 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, where the bright red or blue colors mark high positive or negative correlation. One can see that variables that were initially highly correlated (positively or negativily, Fig. 3) show similar (or in case of negative correlation – 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 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. Such relationship between PC1 and original variables implies that high values of PC1 correspond to ~~describe~~ places~~climate~~ in which seasons are more similar to each other. From the spatial distribution (Fig. ~~6~~5) we can see that PC1 has an east – west gradient implying less distinction between seasons at the seaside. ~~The spatial distribution of PC1 is similar to the spatial patterns of mean start date of winter (see results for Estonia in Jaagus and Ahas (2000)) with higher PC1 values corresponding to later winters.~~ It can be concluded that PC1 reflects continentality of climate, it represents the influence of the Baltic Sea.

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 south. This means that PC2 represents the influence of latitude. ~~However, t~~This pattern is also slightly influenced by geographical features (elevation) and the shape of the coast. ~~The patterns exhibited by PC2 therefore can be expected to be similar to the spatial distribution of phenological events where the 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 beginning of the vegetation of rye (Jaagus and Ahas 2000) or strawberry blooming and harvest (Bethere et al. 2016). In general, higher values of PC2 correspond to earlier phenological processes.~~

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~~ and warm spring~~, or in other words – overall wetter year. PC3 mainly reflects the terrain, i.e. the distribution of elevation.~~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 map of PC3 spatial distribution is similar to the map of average annual precipitation. Interestingly, the precipitation in autumn months (September – October) has a little contribution to PC3 (Table 5).~~

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

Comment [L1]: Šinī nodaļā mēģināju atstāt tikai tīros novērojumus un jebkādas interpretācijas pārnesu uz diskusiju.

~~Conclusions based on spatial pattern and correlation coefficient analysis are summarized in Table 6.~~

3.2 Climate indices for future climate (2071-2100)

Loadings (linear weights) acquired through PCA from 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 period or other geographical locations). It is also important to note that statistics (mean values and variances) from reference data used in data standardization should be applied to other data as well for comparison to be possible. In our case we calculated such climate indices for future climate ~~To calculate indices for future climate~~ (corresponding to period 2071-2100) and analyzed the change in climate~~their~~ patterns. ~~we used loadings acquired from past climate data (see Table 4) and calculated the indices from bias-corrected and standardized data for period 2071-2100.~~ Standardization of variables is shown by Eq. (5) and calculation of climate indices by Eq. (6).

$$\frac{T_k-\bar{T}_k}{\sqrt{\bar{V}(T)}}, \quad \frac{P_k-\bar{P}_k}{\sqrt{\bar{V}(P)}}, \quad k = 1,\dots,12, \qquad (5)$$

where T_k, P_k – temperature and precipitation values for future period, $\bar{T}_k, \bar{P}_k$ - mean temperature and precipitation values for the reference period, $\bar{V}(T), \bar{V}(P)$ – average variance of 12 temperature and precipitation variables for the reference period,

$$Y_i = X_i c_i, \quad i = 1,\dots,24, \qquad (6)$$

where X_i – temperature and precipitation data for future period, c_i – coefficients (loadings) from the reference period, Y_i – climate indices for future period.

It is important to note that Y_i should not be called "principal components" even though they hold similar meaning as prinicipal components from reference data. Y_i are not derived using PCA directly and they do not use eigenvectors from future data.

In Fig. ~~7~~ 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. ~~8~~7. Statistical descriptors, e.g., minimal, maximal and mean value of past and future indices are summarized in Table ~~7~~6. In addition, as we have used the same standardization (subtraction of reference period mean) and climate indices calculation process (loadings from reference period) we can derive conclusions about increase or decrease of these climate indices. However it is important to note that no conclusions can be derived about the value by which the increase/decrease has happened.

All indices have higher values in future climate. This can be interpreted as lower difference between seasons (increase of PC1), an overall warmer climate (increase of PC2) and wetter climate (increase of PC3).

For PC1 it is shown that the values corresponding to coastal regions in reference climate will "move" to the eastern part of Baltic States in the future projections. The expected changes of PC2 are the largest – the maximum values of PC2 for reference climate (in southern Lithuania) are lower than the minimum values for future climate (in central Estonia). Statistics

Formatted: Space Before: 0 pt, After: 0 pt

Formatted: English (U.S.)

in Table ~~7~~ 6 show that the reference range of this index does not overlap with the range of future values. ~~The expected overall increase of PC3 is similar to that of PC1.~~ The climate corresponding to reference values of PC3 in western Lithuania (Zhemaichiai highland) will in future be observable in plateaus in central and north-eastern part of Baltic States.

4 Discussion

Formatted: Normal

The methodology used in this study has been able to reduce 24 climate variables to 3 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 the 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 acquired climate indices can be gained from the literature. The spatial distribution of PC1 is similar to the spatial patterns of 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 the 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 beginning of the vegetation of rye (Jaagus and Ahas 2000) or strawberry blooming and harvest (Bethere et al. 2016). In general, 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 little contribution to PC3 (Table 5).

Conclusions based on spatial pattern and correlation coefficient analysis are summarized in Table ~~6~~7.

~~The methodology can also be applied to the future climate data and therefore the impacts of climate change can be analyzed. Additional analysis is needed for the interpretation of the acquired indices. To some extent such interpretation is provided in this study.~~ The methodology could be further improved to better link acquired indices with phenological processes or seasons by either rotating acquired principal components (Jolliffe 2002) or performing correlation or regression analysis with other variables, such as crop yield (Cai et al. 2013). Another approach that could be used to describe spatial variability of climate in the Baltic States is clustering based on chosen principal component values (Fovell and Fovell 1993, Forsythe et al. 2015).

If variables other than temperature or precipitation are used for the principal components 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

Formatted: Normal

Most of the spatial variability of monthly average temperature and precipitation over the Baltic countries can be represented by three principal components both for past and future climate. These components can be considered as climate indices, where higher values of each index correspond to locations with (1) climate with less distinct seasons, (2) warmer climate, (3) climate with more precipitation. Each component has a distinct spatial pattern. The index related to seasonality exhibits clear east-west (or inland) gradient with less distinct seasonality at seaside (West). The second index (warmer climate) shows north-south gradient with warmer climate in 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 plains and at the seaside. Specific standardization of data allows calculation of such indices also for the future climate. Change in the ~~component values~~climate indices in future implies less distinct seasons, warmer and wetter climate.

Although there are significant change in the magnitude of indices between future and reference period, the change in spatial distribution is relatively small~~For all three indices changes in spatial pattern are minimal~~. For the first and third component regions can be identified where future climate will be similar to the climate currently in other regions.

Acknowlegments

The research was supported by the Latvian State Research Programme “The value and dynamic of Latvia’s ecosystems under changing climate” (EVIDEnT).

The ENSEMBLES data used in this work was funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539) whose support is gratefully acknowledged.

References

The BACC II Author Team ed.: Second assessment of climate change for the Baltic Sea basin, Regional Climate Studies, doi: 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(1-2), 31-41, doi:10.1007/bf00867430, 1997.

Belda, M., Holtanová, E., Halenka, T., and Kalvová, J.: Climate classification revisited: from Köppen to Trewartha, Clim. Res., 59(1), 1-13, doi: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(11), 1231-1262, doi: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(1), 48-58, doi: 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. Climatol., 51(7), 1276-1289, doi:10.1175/jamc-d-11-0168.1, 2012.

Briggs, R.D. and Lemin Jr, R.C.: Delineation of climatic regions in Maine, Can. J. For. Res., 22(6), 801-811, doi: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(4), 719, doi:doi.org/10.1017/s1074070800005228, 2013.

Cattell, R.B.: Factor analysis: an introduction and manual for the psychologist and social scientist, 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, Clim. Chang., 81(1), 329-341, doi: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, Atmósfera, 22(2), 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. Dyn., 43(1-2), 145-161, doi: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. Dyn., 38(7-8), 1359-1373, doi: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(1), doi:10.5194/esd-6-311-2015, 311-326, 2015.

Fovell, R.G. and Fovell, M.Y.C.: Climate zones of the conterminous United States defined using cluster analysis, J. Clim., 6(11), 2103-2135, doi:10.1175/1520-0442, 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(3), 211-224, doi: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(3), 207-219, doi: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(5), 705-720, doi:10.1002/joc.1929 , 2010.

Jolliffe, I.: Principal component analysis, John Wiley & Sons, Ltd, doi:10.1007/b98835, 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(8), 1183-1186, doi:10.1175/1520-0450(1982)021<1183:peitao>2.0.co;2, 1982.

Mahlstein, I., Daniel, J.S., and Solomon, S.: Pace of shifts in climate regions increases with global temperature, Nat. Clim. Chang., 3(8), 739-743, doi:10.1038/nclimate1876, 2013.

Mahlstein, I. and Knutti, R.: Regional climate change patterns identified by cluster analysis, Clim. Dyn., 35(4), 587-600, doi:10.1007/s00382-009-0654-0, 2010.

Malmgren, B.A. and Winter, A.: Climate zonation in Puerto Rico based on principal components analysis and an artificial neural network, J. Clim., 12(4), 977-985, doi:10.1175/1520-0442, 1999.

Netzel, P. and Stepinski, T.: On Using a Clustering Approach for Global Climate Classification, J. Clim., 29(9), pp.3387-3401, doi: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), pp.1-4, doi: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. Discuss., 4(2), 439-473, doi:10.5194/hessd-4-439-2007 , 2007.

Preisendorfer, R. W. and Mobley, C.D.: Principal component analysis in meteorology and oceanography, Amsterdam: Elsevier; doi: , 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(3), 172, doi: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(2), 177-190, doi:10.3354/cr01244 , 2014.

Sennikovs, J., 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(3), 739-751, doi: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, doi: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, 2011.

Zhang, X. and Yan, X.: Spatiotemporal change in geographical distribution of global climate types in the context of climate warming, Clim. Dyn., 43(3-4), 595-605, doi:10.1007/s00382-013-2019-y, 2014.

Figures and Tables

Figures

(FIGURE REMOVED)

~~Figure 1: Locations of meteorological stations used for bias-correction of RCM data.~~

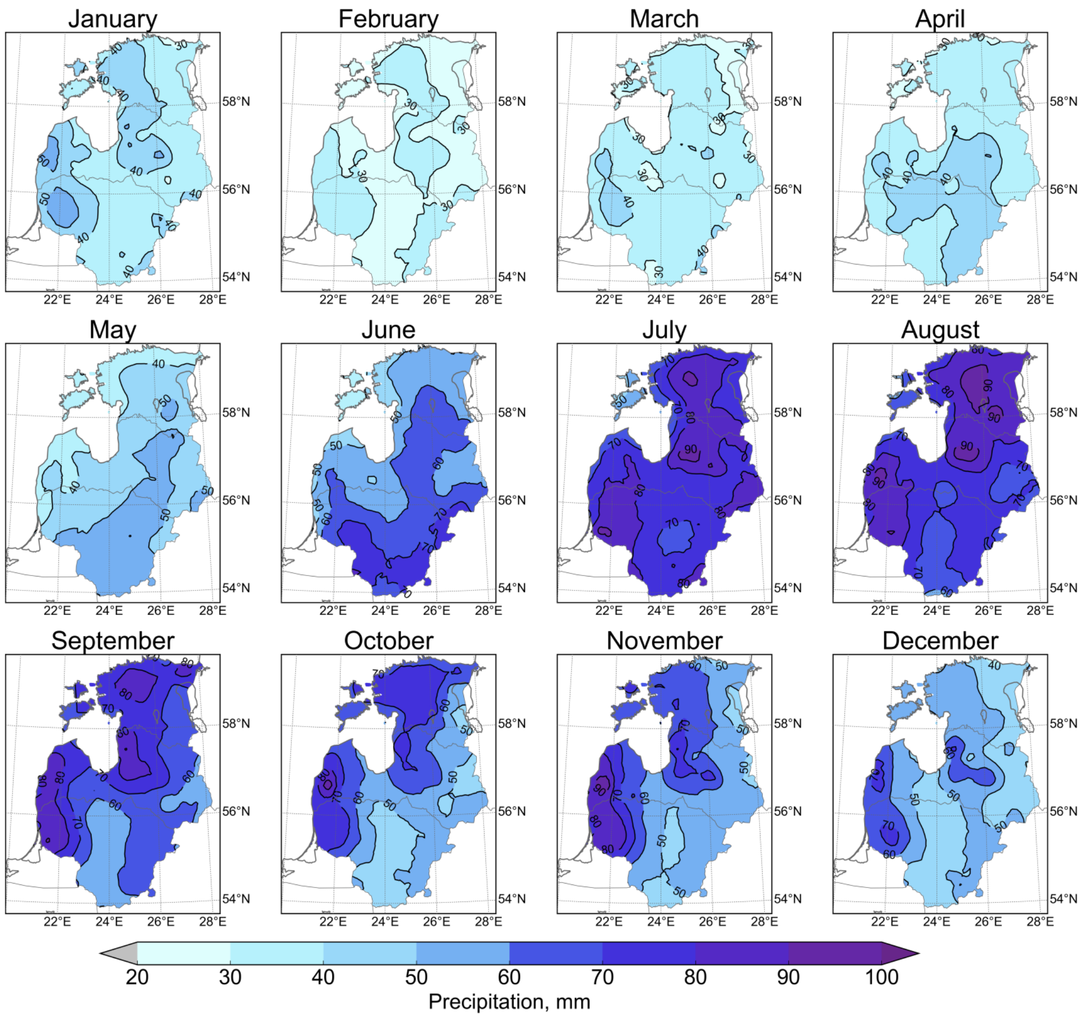


Figure ~~2~~1: Monthly precipitation 1961-1990, bias-corrected median of RCM ensemble.

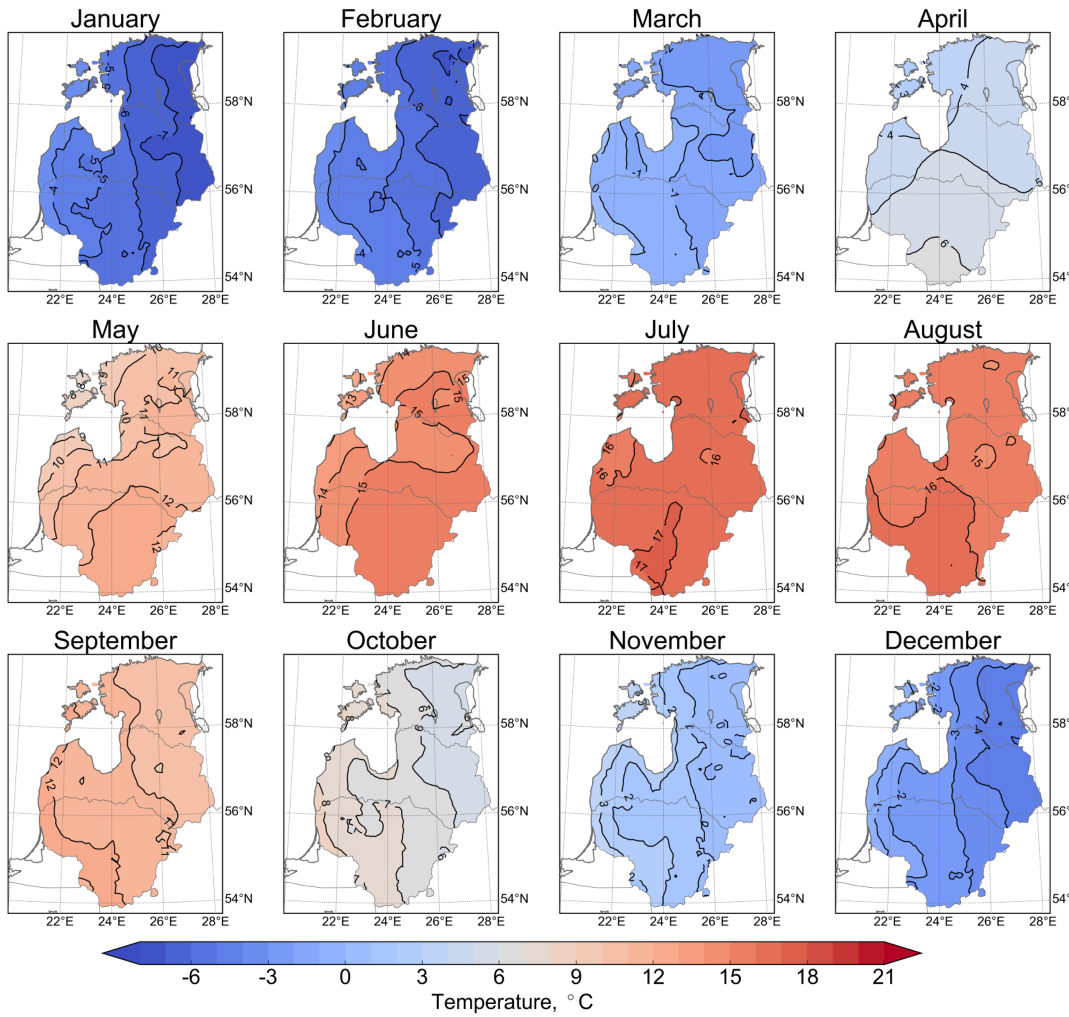


Figure ~~3~~2: Monthly average temperature 1961-1990, bias-corrected median of RCM ensemble.

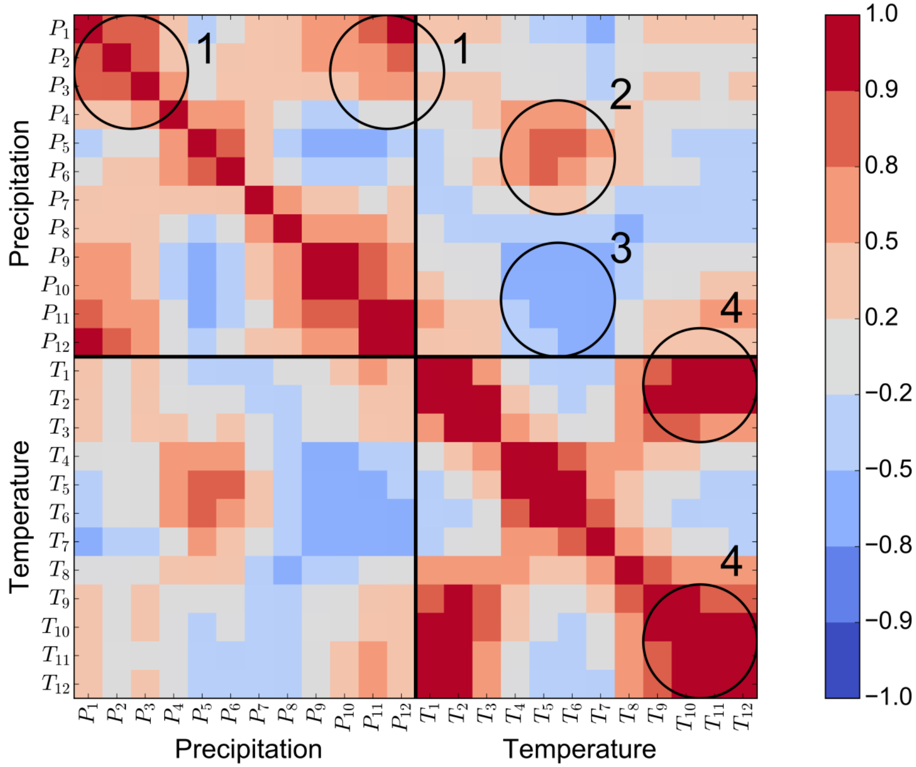


Figure 43: Temperature-precipitation correlation matrix, bias-corrected data. Marked and numbered features show especially high absolute correlation: 1- strong correlation between precipitation in winter months; 2- strong correlation between precipitation and temperature in spring months; 3- strong negative correlation between precipitation in autumn and spring temperature; 4- strong correlation between temperatures of autumn and winter months.

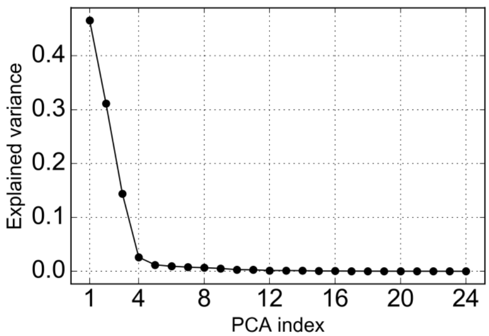


Figure ~~5~~4: Scree plot (explained variance of each principal component), calculated for reference (1961-1990) climate.

(FIGURE REPLACED)

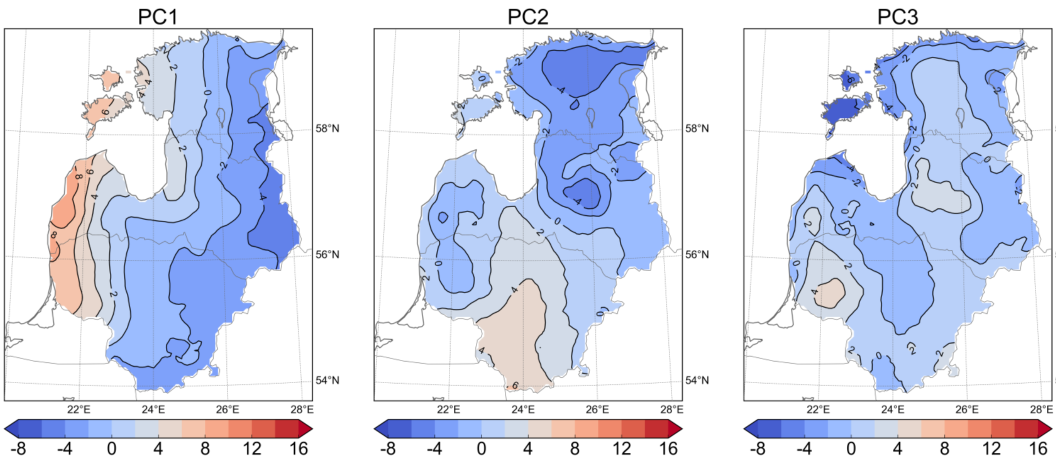


Figure ~~6~~5: Spatial pattern of first three principal components based on monthly temperature and precipitation data for years 1961-1990.

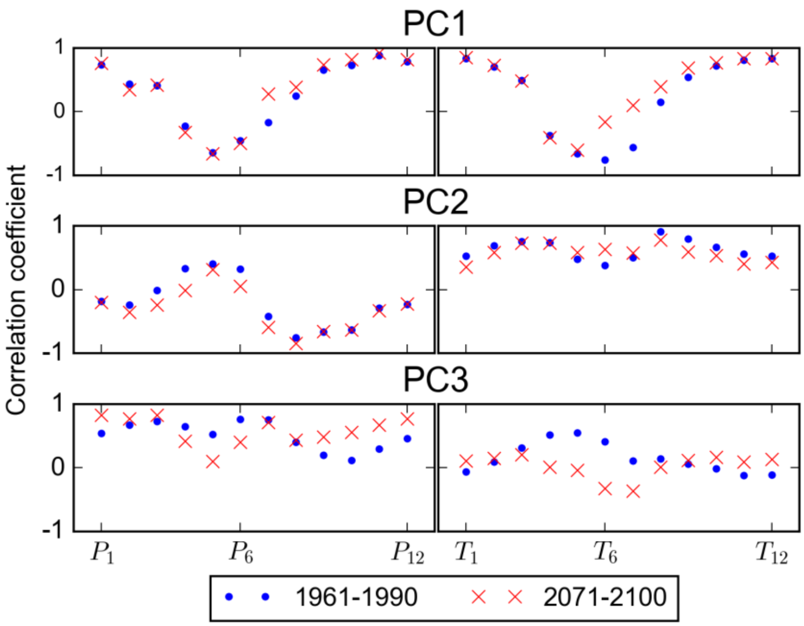


Figure 6~~7~~: Correlation coefficients between indices (principal components) and initial variables for reference and future climate.

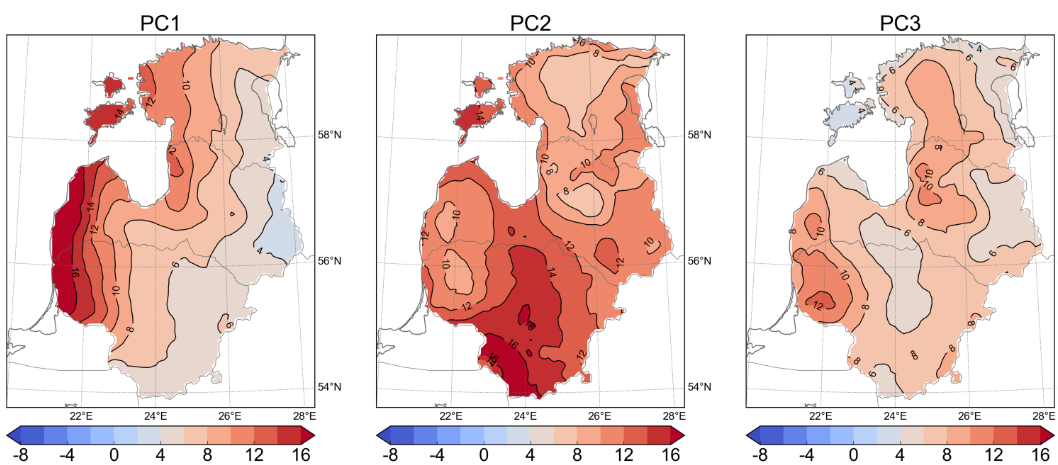


Figure 7~~8~~: Climate indices (based on principal components from 1961-1990) for years 2071-2100.

Tables

Table 1: List of the Regional Climate Model (RCM) ensemble members used (ENSEMBLES), showing the originating institution, the name of RCM, the driving General Circulation Model (GCM). For explanation of abbreviations see van der Linden & 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 sens.)
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

Table 2: Variances of climate variables before and after standardization for 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

Table 3: Variances of climate variables before and after standardization for 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

Table 4: Explained variance and loadings of first 3 principal components, calculated from temperature and precipitation data for 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	

Table 5: Correlation coefficients between principal components and standardized initial data for 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

~~Table 6: Description and interpretation of climate indices base on PCA.~~

~~Name~~	~~High values correspond to~~	~~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~~

Table 6~~7~~: 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

Table ~~6~~7: Description and interpretation of climate indices base 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