Referee comments to the manuscript

Climate indices for Baltic States from principal component analysis

by Liga Bethere, Juris Sennikovs and Uldis Bethers

The authors have presented an interesting article concerning climate indices for the Baltic region. Principal component analysis (PCA) was applied for monthly mean air temperature and monthly precipitation, which were derived from the ensemble of bias-corrected regional climate models output for the territory of Estonia, Latvia and Lithuania in 1951-1990. Three main climatic indices were defined and described. The indices were calculated for the period 2071-2100 using the model outputs.

I suggest that the manuscript is worth of publishing but it needs major revision. There are many aspects, which are not sufficiently explained. The interpretation of the results could be improved. My concrete remarks and suggestions are the following.

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

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

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?

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

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.

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.

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.

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

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

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.

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

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?

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.

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

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.

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.

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.

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.

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.

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.