Earth Syst. Dynam. Discuss.,
doi:10.5194/esd-2017-34-RC2, 2017

Earth System Dynamics Discussions
Open Access
EGU

Interactive comment on “Climate indices for Baltic States from principal component analysis” *by* Liga Bethere et al.

Anonymous Referee #2

Received and published: 19 May 2017

In this paper the authors use PCA to derive climate indices for the Baltic states. Rather than use observations directly, the authors incorporate RCM data into the analysis to investigate the effect of climate change on their indices.

The abstract states that the authors are interested in the spatial structure of climate in the study region. Given this aim, the authors chose a suitable method. However, the interpretation of the results is often (in my view) incorrect and the discussion touches repeatedly on the importance of the seasonal cycle. Given this, I suggest the authors modify their aim and conduct a more conventional PCA analysis. I will return to this point below.

As an example of the problems in interpreting the results:

page 3, line 20, interpretation of Figure 4. "more precipitation in, e.g., December is

Printer-friendly version

Discussion paper

clearly linked to more precipitation in January. Thus, entire winters are either dry or humid"

This reads as if you are implying some temporal correlation (a dry December is linked to a dry January following), whereas what you really mean is that the spatial precipitation patterns are similar in these months.

"Such relationship between PC1 and original variables implies that high values of PC1 describe climate in which seasons are more similar to each other. "

I do not agree that PC1 represents the magnitude of the seasonal cycle. The seasonal cycle was removed from the 1961-1990 dataset during the normalization. The distinct annual cycle in the PC1 loadings is because the main spatial variation (distance from the coast/continentality) varies seasonally. However, this annual cycle is NOT the same as the seasonal cycle of temperature and precipitation, as per the “normal” definition. For example, the temperature loadings for PC1 (Table 4) have an annual cycle with max in Dec/Jan and a min in May. The seasonal temperature cycle (approximately from Fig 3) has a min in Jan/Feb but a max in July. 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! This, for me, is actually the most interesting result from the paper! I can offer no physical explanation for this, and I hope you receive some ideas from this discussion forum! Perhaps you could plot the PC1 loadings on the same graph as the normalized seasonal cycle to demonstrate the result.

Given the above interpretation – that PC1 represents the annual cycle of the distance-to-the-coast effect - I do not believe that the 1961-90 means and variances can be used to normalize the 2071-2100 data. To illustrate, consider a hypothetical climate change that results only in a change in the temperature seasonal cycle (that is, the change in all temperature grid cells for each month is a constant, but the changes are different for each month). That is, there is no change in the spatial structure in the data. But because the temperature season cycle is not orthogonal to the temperature distance-to-the-coast annual cycle, this change would cause "PC1" 2017-2100 (as calculated in the paper) to differ from PC1, 1961-90. I do not believe interpreting the difference as a change in the continentallity is valid if there is no change in the spatial structure of the dataset!

It is difficult for me to conceptualize the value of a purely spatial climate index for a small region, and the numerous references to the importance of seasonality (mean start date of winter, phenological events, spring floods) suggests that the authors have similar difficulties.

I would suggest that the results would be much easier to interpret if the precipitation data were simply scaled (as one distribution, all months and grid cells) to have the same variance as the temperature data, then the combined dataset normalized (again as one distribution). The first set of loadings will then be the average annual cycle for precipitation and temperature, which is the most important component of any climate index from the point of view of applications. Scaling to the future data in the same way will allow a direct estimate of changes in seasonality.

If the current scaling is retained, the discussion must focus on spatial aspects of the PCs, and not the temporal aspects of the loadings. How do the PCs correlate with distance-to-the-coast? To elevation, slope and aspect?

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?)

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

Interactive comment on Earth Syst. Dynam. Discuss., doi:10.5194/esd-2017-34, 2017.