

Interactive comment on “Uncovering the shortcomings of a weather typing based statistical downscaling method” by Els Van Uytven et al.

Mohammad Sohrabi

msohrabi2@ucmerced.edu

Received and published: 24 May 2019

This study aims at evaluating assumptions of a weather typing (WT) based statistical downscaling method (SDM) for precipitation and river peak flows in Belgium. The results of such studies provide an assessment of end user needs in choosing a right downscaling methods out of many available methods in terms of the cons and pros of each method and the intended use of the results. However, the current study for uncovering the shortcomings of a downscaling method has several serious shortcomings itself as listed below:

1. Validity of the study

The results of this study showed that the synoptic changes (WT occurrence changes)

C1

contributed only 20% of the total change in daily precipitation and the change is mostly (80%) explained by other processes including the thermo-dynamical processes. It obviously indicates that a weather typing based statistical downscaling method shouldn't be used in the region for the downscaling of precipitation which is mainly originated from local moisture. So, what would be the point for evaluating a downscaling method which cannot be used in the region?

2. Novelty of the study

The statistical downscaling method (SDM) was taken from the literature without any modification: SD-B-7 method from Willems and Vrac (2011). The evaluation of statistical downscaling methods is not also new, although it is mentioned in the second line of the abstract that “Each statistical downscaling method (SDM) has strengths and limitations, but those are rarely evaluated”. Nine more comprehensive studies were mentioned in P3L15 of this paper (some of them from EU COST Action VALUE project for the evaluation of statistical downscaling methods) along with several unmentioned. A clear objective was not defined for this study. While in the abstract, only extreme precipitation and river peak flows were stated, the majority of the paper is about daily precipitation and not extremes. As an example, it was claimed (P6L12-15) that compared to the previous study by Brisson et al. (2011) across different stations in Belgium, the current study focuses on the extreme precipitation amounts. However, the presented results are more related to winter precipitation accumulation and percentage of wet days per WT (Figures 2 and 3). Overall, there is not a consistent storyline in the paper. It appears that the paper is a combination of several small studies (leftover results actually) and then fitting a statistical downscaling to them.

3. Statistical analysis and extreme event definition

Another major issue in the paper is that the results for extremes are based on a limited sample size. The extremes were separated based on 4 seasons and 11 weather types. How extremes are the selected extremes for each season and weather type? What

C2

threshold was used for defining extremes? Apparently, the return period of 0.1 year was chosen as the threshold. The question would be whether precipitation and streamflow that occur on average every month is really considered extreme in hydrology. Due to a small sample size after the separation per season and per weather type, even an extreme precipitation of a 10-year return period amounts 0.5 mm/hr (Figure 5).

4. Justification of the selected methods

a) Why was the Lamb weather classification used in this study, while k-means clustering is regarded as one of the best-performing classification schemes over western Europe (e.g., Beck and Philipp, 2010; Casado et al., 2010; Garcia-Valero et al., 2012; Broderick and Fealy, 2015).

b) De Niel et al. (2018b) identified a minor uncertainty contribution by the hydrological models in the peak flow changes. Among the tested hydrological models in that study, why NAM was selected for the current study?

c) Why were these three reanalysis datasets selected for this study? Why didn't they use the E-OBS observations (1950-2018) which has similar data coverage to the ERA40 reanalysis dataset (1948-2002)?

d) Several downscaling methods were developed and evaluated by Willems and Vrac (2011). Why was the SD-B-7 method selected for this study?

5. Weather types

a) The climate model results for WTs are largely biased. Although the bias was reported in the text regarding the mean scenario of the climate models, the difference goes up to 20% for the anti-cyclonic (A) WT and 30% for west (W) WT: none of the climate models can even reach the frequency of A WT estimated by the reanalysis datasets. Considering these large biases, how reliable would be the downscaling results based on these WTs? The climate models with a coarse resolution are expected to reasonably simulate these large scale patterns, and so what might be the main rea-

C3

son for such a bias in climate model results? I would be interesting to investigate why the GCMs have the largest uncertainty for the W WT as the main large scale driver of winter precipitation over western Europe?

b) What is the driver for the undefined weather type or the atmospheric state characterized by a weak flow? A sensitivity test of unclassified days on grid sizes and resolutions by Demuzere et al. (2009) showed that the number of these days decreases with grid resolution. Was that the case for this study as well? And generally, how will the different resolutions of reanalysis data from 2° to 5° explain the discrepancy between reanalysis-based WTs?

c) A separate set of WTs was produced for each season in the current study. What will be the influence of the seasonal cycle on the classification produced as the MSLP fields are clustered?

6. Statistical downscaling by analogues

The CMIP5 GCMs provide data at a daily time scale. Were daily precipitation data from the GCM scenario period corresponded to observed sub-daily precipitation? If so, how are the results influenced by the difference in the time scale? Were the climate change signals assumed to be time scale dependent?

7. Scaling by the Clausius-Clapeyron relation

a) Clausius-Clapeyron relation assumes that extreme precipitation amounts are controlled by local moisture availability. How local is moisture availability? The developed extreme precipitation-temperature scaling relations for central Belgium were used for river peak flow simulations in a catchment in the northeast of Belgium. How representative would be the scaling relations developed in central Belgium for northeast Belgium considering the local moisture availability assumption?

b) Dry-bulb temperature was used here for developing scaling relations, whereas several recent studies (e.g., Wasko et al., 2018) recommended to use dew point tempera-

C4

ture than dry-bulb temperature for Clausius-Clapeyron relation, as it is a better measure of precipitation changes because of increases in the moisture holding capacity of the atmosphere (Lenderink et al., 2011).

8. Evaluation of greenhouse gas scenario assumption

To evaluate the greenhouse gas scenario assumption of the SDM, changes in the WT occurrences and average daily temperature as a function of the four RCPs were analyzed. However, this assumption might be tested for precipitation which is statistically downscaled in this study. Besides, the increase of the change in air temperature with greenhouse gas emissions is trivial. What is the relation of changes in warm extremes (with return periods ranging between 0.1 and 10 years) (Figures 7 and 8) to the downscaling of precipitation performed in this study? Though irrelevant, for warm extremes the changes in maximum temperature should be analyzed instead of mean daily temperature. P7L9-10: "To check whether the predictor simulation results are adequate and accurate, a comparison is made between the climate model simulated and observed daily average temperature statistics". Is predictor warm extremes? It seems that the full range of temperature was used for the scaling (Figure 6).

9. Evaluation of the stationarity assumption

For the evaluation of the stationarity assumption, the extreme precipitation per weather type was compared for different sub-periods of 10 years length between 1901 and 1991. Use of a 10-year sub-period for this purpose is questionable as it is far smaller than a natural climate cycle, and hence the results are greatly influenced by natural climate variability. Isn't trend analysis a more robust approach for testing the stationarity? How large is the uncertainty in the presented results? In my view, a random selection of a dry weather type or and performing the same analysis or doing the would reveal the reliability of this results? Would performing the decadal analysis of extreme precipitation without considering the WTs lead to similar results? This is because based on the results of this study, weather types shouldn't be related to precipitation formation

C5

in the region (see comment 1).

10. Evaluation of the method specific assumptions

The evaluation of the SDM method specific assumptions was only performed for winter as the peak flows in the selected catchment mainly occur in the winter (P5L12-14). As mentioned in P4L30, "Application of this scaling rate to precipitation intensities is valid assuming that extreme precipitation amounts are controlled by local moisture availability and are not influenced by large scale atmospheric circulation patterns." However, the influence of the large scale circulations on winter precipitation in western Europe is well documented in the literature.

11. Interpretation and discussion of the results

a) The results were interpreted and discussed in a way that the authors expected the results be. For example, in P12L20-25, the authors attribute the difference between their results and those of Otero et al. (2018) for changes in the anti-cyclonic WT to considered climate model ensemble, the location of the 16 points grid for the WT classification system and, the reference and scenario periods. I am wondering why these differences between the two studies (this study and Otero et al., 2018) are only important for the changes in the anti-cyclonic WT and not for the changes in cyclonic, west and southwest WTs! Another example is in P15L28-31, where the authors mention only the results for RCP4.5 and RCP8.5 and not all RCPs to show that changes in the WT occurrences and precipitation are magnified under increasing greenhouse gas scenarios. Looking at the results, changes in the occurrence of W WTs are far smaller for RCP8.5 compared to RCP6.0 (P12L5). Also, there is not a clear pattern for the changes in cyclonic (C) WTs where changes are equal to 5% for RCP 2.6, - 3% for RCP 4.5, -6% for RCP 6.0 and -5% for RCP 8.5. It was speculated in P12L8-10 that these discontinuities in the uni-directionality of the changes may be explained by the smaller ensemble size for RCP 6.0 compared to the other RCPs, and/or by the different RCP sub-ensemble compositions. This issue can be easily checked by selecting

C6

the same GCMs for different RCPs. Why isn't this an issue for temperature changes (P12L28-30).

b) In several places in the text, internal variability was argued as the reason for unexplained behaviors in the results: for example, internal variability as the reason for the large bias of climate models for WT simulations and also internal variability among the climate models as the reason for discontinuities in the uni-directionality of the changes with greenhouse gas scenarios. These statements might be supported by the results or the literature.

c) The results for the changes in the frequency of WTs showed that the frequency of the wet WTs will increase under climate change and the frequency of the dry WTs will decrease. It is worth discussing what physical explanations are behind the increasing frequency of westerlies and the decreasing frequencies of easterlies under climate change. How might global warming decrease the frequencies of cyclonic and anti-cyclonic WTs? The response to these important questions would be beneficial for improving weather typing based SDMs.

d) "Stationary is dead" is now a fact for hydrologists. What would be the contribution of this part of analysis to the existing knowledge? Rather than discussing the natural climate anomalies for this single location in western Europe in section 4.5, it would be more useful to discuss the drivers for such anomalies, as several global studies regarding these historical natural anomalies of the climate system have already been published. P13L21-22: "the 10 minutes precipitation amounts with a 1 year return period measures 6mm/h for the negative anomaly, whereas it is 14mm/h for the positive anomaly". What might be the reason for the positive and negative anomalies in W WT and the related extreme precipitation? P13L20 & P13L25: "The difference between the positive and negative precipitation anomaly is especially visible for the W WTs and this for all aggregation levels between 10 minutes to 1 day. For the C WTs, no differences appear between the amounts for the positive and negative anomaly". Why is W WT-based extreme precipitation timescale-dependent, but not the C WT-based extreme

C7

precipitation?

12. Reanalysis data uncertainty

The uncertainty related to the choice of reanalysis data for WTs was considered. Given that the reliability of reanalysis products sharply decreases back in time (Ferguson and Villarini, 2013; Krueger et al., 2013) due to assimilating sparse observations and starting from a more uncertain initial state (Delaygue et al., 2019), a larger uncertainty of reanalysis data is expected for earlier years of the study period. Was the uncertainty calculated for the entire analysis period in this study? Does the uncertainty decrease as time progresses from far past to near past?

Specific comments P5L19: The evapotranspiration data are not daily measured data for 100 years, are they? Is it reference evapotranspiration or potential evapotranspiration? If the former is the case, with grass or alfalfa as the reference crop?

P7L15: 33 unique control runs were used in this study. I think the number of independent model runs should be lower than that! How was the dependency between climate models investigated?

P7L16-17: The authors found that the choice of the reference period (control period) influences the evaluation of the perfect prognosis assumption. It would be interesting to present the results of the sensitivity analysis to the control period in the Supplementary Information.

Figure 12: Does grey area show the 5th-95th percentile range of climate change uncertainty?

Table 2: Were the same GCMs used for all climate variables studied?

Used references not present in the reference list of the paper

1. Beck, C., & Philipp, A. (2010). Evaluation and comparison of circulation type classifications for the European domain. Physics and Chemistry of the Earth, Parts A/B/C,

C8

35(9-12), 374-387. 2. Broderick, C., & Fealy, R. (2015). An analysis of the synoptic and climatological applicability of circulation type classifications for Ireland. *International Journal of Climatology*, 35(4), 481-505. 3. Casado, M. J., Pastor, M. A., & Doblas-Reyes, F. J. (2010). Links between circulation types and precipitation over Spain. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(9-12), 437-447. 4. Delaygue, G., Brönnimann, S., Jones, P. D., Blanchet, J., & Schwander, M. (2018). Reconstruction of Lamb weather type series back to the eighteenth century. *Climate Dynamics*, 1-18. 5. Ferguson, C. R., & Villarini, G. (2014). An evaluation of the statistical homogeneity of the Twentieth Century Reanalysis. *Climate Dynamics*, 42(11-12), 2841-2866. 6. García-Valero, J. A., Montavez, J. P., Jerez, S., Gómez-Navarro, J. J., Lorente-Plazas, R., & Jiménez-Guerrero, P. (2012). A seasonal study of the atmospheric dynamics over the Iberian Peninsula based on circulation types. *Theoretical and applied climatology*, 110(1-2), 291-310. 7. Krueger, O., Schenk, F., Feser, F., & Weisse, R. (2013). Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. *Journal of Climate*, 26(3), 868-874. 8. Lenderink, G., Mok, H. Y., Lee, T. C., & Van Oldenborgh, G. J. (2011). Scaling and trends of hourly precipitation extremes in two different climate zones—Hong Kong and the Netherlands. *Hydrology and Earth System Sciences*, 15(9), 3033-3041. 9. Wasko, C., Lu, W. T., & Mehrotra, R. (2018). Relationship of extreme precipitation, dry-bulb temperature, and dew point temperature across Australia. *Environmental Research Letters*, 13(7), 074031.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2019-40>, 2019.