

Comments on “Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe” by M. A. Sunyer et al.

General comments

We would like to thank both reviewers for providing clear and relevant reviews and suggestions for improvement. We have tried our best to incorporate all the comments and suggestions from all the reviewers, which we think have improved the quality of the paper.

In addition to the comments from the reviewers, and following the suggestions from the editors, we have extended our literature review and evaluated our results in more detail in relation to the existing literature, see changes in e.g.: **second paragraph page 4, first paragraph page 5, page 18 line 13 – 31, page 21 line 6 – 10, page 25 line 9 – 20**. We have left the acronyms for the catchments in the manuscript. We have not created an appendix with a list of the acronyms as they are specified in Table 1, together with the main characteristics of the catchments. The marked-up manuscript version is included after the comments of the referees, all the changes have been highlighted in yellow.

Anonymous Referee #1

Received and published: 10 July 2014

The authors apply different bias-correction/downscaling (BC/DS) methods to RCM simulated daily precipitation time series in 11 European catchments. They evaluate how the methods differ both with respect to the agreement with observations in the control period and with respect to future changes, in both cases with focus on extreme values of duration between 1 day and 1 month. Overall limited differences are found with weak dependence on e.g. location and duration. Precipitation BC/DS is generally a key activity in hydrological climate change impact studies and evaluation/comparison of methods is an important activity. The experiment is very comprehensive, spanning a wide range of climate projections, methods and catchments. The outcome must be worth sharing in the scientific community, but I think substantial revision of this manuscript is needed before publication, mainly for the reasons discussed in the following.

1. The overall conclusion is essentially that there is no best method but we must use many, and on an as large an ensemble of RCM-projections as possible. In reality, however, it is extremely rare to have such resources but the impact study must be limited to one BC/DS method applied to a small RCM-ensemble (or even just one projection, it is not unusual). Thus I think that a key objective of this kind of study must be to provide advice and recommendations for real-world, limited-resource impact studies. If precipitation extremes are the key focus, what should we do? With all these results and all these (prominent) authors it must be possible to provide more useful knowledge, conclusions and advice than what is currently the case, only “highlight the need for considering an ensemble” is not of enough help. For example different methods are to various degree prone to different problems like (1) cannot be applied/create unrealistic extreme values under some circumstances, (2) increases the bias in precipitation extremes under some circumstances, (3) modifies the climate change signal with respect to changes in extreme precipitation under some circumstances, (4, 5: : :). Further different methods are more or less

prone to deviate from the rest in other aspects. A systematic review of this kind of key issues would provide very useful information.

- We have now extended the discussion on the selection of downscaling methods. It is difficult to point to a specific best method as it depends on several factors and the results might be different depending on the application, but we have pointed out some issues that arise from the use of some methods (such as BCM, BCMV, CFM,...), which implies that they should not be selected for some specific applications (see **page 21 lines 21 – 26**). Throughout the paper we have also extended the discussion regarding the ability of the models to correct the RCM outputs (**page 23 lines 2 – 12, page 25 lines 23 – 24**) as well as preserve the climate change signal (see **page 17 lines 13 – 23 and page 21 lines 11 – 21**) from the RCMs (as recommended in some other comments below). This helps to provide information on which methods should and should not be selected and why. We think that it is important to highlight the need of using (when possible) an ensemble of methods to be able to assess the uncertainty in extreme precipitation projections. We have described the main characteristics that the methods included in the ensemble should cover (different underlying assumptions, different RCM outputs, preserving climate change signal, ...) (see **also page 27 line 13 – page 28 line 4 and page 28 lines 11 – 15**).

2. As I understand it the authors have calibrated the methods on the full set of reference data available and then applied to the same set of data (as well as the future-period data). However, I think this type of study needs to also include some kind of cross-validation analysis for historical periods, i.e. calibrate on one period and verify for another. This may not be possible for all catchments but in many there are some 50 years of data so you could split equally, divide 30/20 or something else. I think this kind of analysis is crucial for assessing the uncertainty when applying the methods to future periods.

- We agree that the validation of these methods is crucial. However, the proper validation of statistical downscaling methods is difficult and should be carried out with care. As recommended in several studies (e.g. Refsgaard et al. 2014; Teutschbein and Seibert, 2013), it should be done using data that has different properties to be able to assess whether the downscaling methods can be used to project climate changes. It would be possible to carry out validation analyses with the data available, but if the observational data do not show a pronounced change in extremes, then the results of the validation analyses would be questionable. The validation of statistical downscaling methods is very relevant and needed but it would require an almost new study in itself. A range of recent studies (e.g. Räisänen and Rätty 2013; Refsgaard et al. 2014; VALUE cost Action <http://value-cost.eu/>) focus exclusively on methods and techniques to properly validate statistical downscaling methods. We consider that further research continuing this study should focus on the validation of the methods used here. This discussion regarding the challenges in the validation of statistical downscaling methods has been added to the manuscript (see **page 4 lines 3 – 6, page 22 lines 19 – 31, and page 24 lines 9 – 19**).

Räisänen, J. and Räty, O.: Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations, *Clim. Dyn.*, 41, 1553–1568, doi:10.1007/s00382-012-1515-9, 2013.

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Trolle, D., Willems, P. and Christensen, J. H.: A framework for testing the ability of models to project climate change and its impacts, *Clim. Change*, 122, 271–282, doi:10.1007/s10584-013-0990-2, 2014.

Teutschbein, C. and Seibert J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions?, *Hydrol. Earth Syst. Sci.*, 17, 5061–5077, doi: 10.5194/hess-17-5061-2013, 2013.

3. I find the presentation of results hard to follow. Results from different periods focusing on different aspects are mixed under the headings “all catchments” and “selected catchments”. I think the paper would benefit from being structured more like the bullets 1-4 on l 13-25, p 6181.

- We agree that the headings used for the different sections are not very helpful for the reader to identify what is discussed in each section. The sections are organized as the 1-4 bullet points in page 13. We have renamed the sections in order to link their name with the bullet points. For example, the section 4.1 called “Statistical Downscaling methods” has been renamed to “Comparison of the downscaled time series for the control and future periods” so it matches the comparison in bullet point 1. Similar changes have been applied to the other headings in the results section.

4. Related, I think the paper would benefit substantially from more distinct objectives. Now they are rather vaguely and incompletely formulated like “comparing BC/DS methods” or “assess the changes in extreme precipitation”. It would be better to formulate some distinct hypotheses to investigate. Focus on apparent current knowledge gaps that this study can help filling.

- We agree that the objectives were not clearly stated. We have now described more clearly the key objectives of this study and how they relate with the current knowledge gaps. The key objectives of this study are: to identify if there are general similarities and/or dissimilarities between the statistical downscaling methods; to assess whether there are common trends regarding changes in extreme precipitation over Europe; to assess the main sources that lead to the variability in the results. **See changes in the text in page 5 line 10 and lines 18 – 23.**

The Results section is largely a textual description of the tables and figures with little interpretation or explanation or speculation of the reasons behind the findings. Sometimes it is speculated, but in those cases they could often have gone back to the results and for verification. See further specific comments below.

Specific comments:

5. 20, 6169: It sounds puzzling that bias correction improves the agreement with observations only “in most cases”, clarify that is concerns the extremes.

- The fact that some of the bias correction methods do not lead to an improvement compared to the RCMs is now discussed in more detail. See discussion in comment **22**.

6. 13-23, 6171: You need to indicate also what relevant knowledge that was found in these studies.

- The main conclusions regarding the influence of different sources of uncertainty from these studies are now discussed in the text. These are: Bürger et al., (2013) concluded that the main influences on the overall results for different extreme indices were the downscaling method followed by the climate model used. Sunyer et al. (2012) and Hanel et al. (2013) highlighted that the influence of the statistical downscaling method used on the variation of the results is more pronounced in the case of extreme events (extreme precipitation in the case of Sunyer et al. (2012) and droughts in the case of Hanel et al. (2013)). Wilby and Harris (2006) concluded that in the case of low flows the main sources of variation are the statistical downscaling methods (SDM) and climate models used. Lawrence and Haddeland (2011) showed that in rainfall dominated catchments, the uncertainty arising from the hydrological parameters was more significant than other sources. But in snow dominated catchments, climate scenarios and SDMs were the main source of uncertainty. This discussion has been added in the manuscript in **page 4 lines 12 – 23, 26 – 28, and 29 – (page 5) 9**.

7. 24, 6171: Is not comparing BC/DS methods rather the main focus?

- Yes, the comparison of statistical downscaling methods is one of the main objectives of this study. In addition, this study also addresses the main sources of variability in the results as well as comparing the magnitude and direction of the changes in eleven European catchments. See the comment **4** regarding the need for more distinct objectives. The different objectives of the study have been stated more clearly in the manuscript.

8. 3-6, 6173: Was the gridded data not based on observations?, if so how were they derived?

- The gridded data was based on observations. The information regarding the methods used to obtain the gridded data is given in the references specified in Table 1. It has been clarified in the text that the gridded data is based **see page 6 line 20** on observations from station data,.

9. 2.1, 2.2: What was the volume resolution in the observations?, 0.1 mm?, was the same cut-off used for the RCM data for consistency?

- The cut-off value in the observations varies depending on each catchment. The cut-off values were not applied to the RCMs as it is not relevant for the study of extreme precipitation. Nonetheless, several downscaling methods include a threshold of wet/dry days which addresses the issue of the cut-off values (in addition to accounting for changes in the frequency of wet/dry days in a future climate). The different cut-off values of the

observations and the fact that they are not used in the RCMs is now discussed in the text (see page 6 line 21 – 25).

10. 5, 6174: Rummukainen is misspelled and not in reference list, check carefully before submitting, should not be my task.

- Corrected.

11. 10-12, 6175: I think harmonisation would have been much better, why not used?

- This paper is a result of a coordinated effort where different research groups contributed with the different methods that they use. We agree that harmonisation would have been nicer but the methods were applied as previously used in other studies by the group. The reason for the approach used is now mentioned in the manuscript (see page 8 line 20 – 23).

12. 22-23, 6176: What if there are more wet days in OBS than in RCM?

- This is not the case for any of the RCMs when the bias correction is applied to the full time series, without distinguishing months or seasons, as has been done in this application.

13. 19, 6177 – 5, 6178: As acknowledged it seems questionable to train on RCM-ERA and then apply on RCM-GCM without any correction. Can you provide some information on the additional error introduced?

- With respect to the downscaling, the XDS model is optimally trained given the available RCM data. With respect to the RCM data, however, better RCM realizations could be achieved by a second data assimilation with the RCM-ERA-40 runs (beyond the data assimilation which is already done for the ERA-40 reanalysis) in a similar way as done e.g. in the North American Regional Reanalysis (NARR) (Mesinger et al., 2010). The additional error introduced when applying XDS to RCM-GCM simulations cannot be quantified. For clarification we have added the following sentence to **page 10 line 20 - 23**: “A second data assimilation with the RCM-ERA-40 runs (beyond the data assimilation which is already done for the ERA-40 reanalysis) would overcome this problem to some degree. However, such runs are not available for the RCMs accessible from the ENSEMBLES archive.”

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D. and Shi, W.: North American Regional Reanalysis, 2010.

14. 9-12, 6179: The intervals are very small (and different from BCQM). It is well known that there may be an enormous variation in the very highest quantiles (above 99 or so), I think it needs to

be demonstrated that using 0.0005 works well, instead of smoothing out the fluctuations a bit with larger intervals.

- The probability intervals are estimated using linear interpolation from the empirical quantiles, so as in the case of CFQP and BCQM the fluctuations in the CFs are due to the fluctuations in the empirical quantiles. The possibility of large fluctuations in the changes estimated using empirical quantiles is a disadvantage of using empirical methods as used in this study, but empirical methods have the advantage that there is no need to fit a distribution to the data. In methods where distributions are fitted changes would be smoothed; however, with a risk to over-smooth the changes and introduce a bias in the estimated quantiles. It should also be noted that in most catchments the results of CFQM and CFQP are virtually the same, and CFQP uses different intervals than CFQM (but also estimated from the empirical quantiles). This discussion on the problem of fluctuations in the CF has been added to the text, **see page 12 line 23 – 26.**

15. 15, 6185: What is a “threshold return level”?

- Corrected to only “thresholds”.

16. 20-27, 6185: Assessing to which degree BC/CS modifies the CC signal is important and needs more attention. More analysis and interpretation is needed. What is the change increased after BC/DS? What is the reason for the regional changes? Dig in the results.

- We agree that it is important to assess whether the difference in the change projected for extreme precipitation using the downscaled time series differs from the change projected by the uncorrected RCMs. This has now been discussed in more detail using the overall results for all the catchments (results in Figure 3) and in detail for each downscaling method for three of the catchments (results in Figure 4). Overall the change in signal introduced by the downscaled series is not significant compared to the change projected by the RCMs, but some SDMs (CFM and in some cases XDS, BCM, and BCMV) tend to lead to different values than the ones obtained from the RCMs. These methods do not specifically correct or take into account the changes in extremes which might lead to differences in CC signal. On the other hand, the methods that are expected to better account for changes in extremes tend to show more similar results to the ones obtained for the RCMs, especially the BCQM method. It is now discussed in the paper (in the description of the methods and further emphasised in the discussion of results) how the different methods are suited for handling changes in extremes. There are not specific regional differences in the change of the CC signal from the downscaled series and the RCM series, except the fact that larger differences are obtained for the catchments in Turkey and Cyprus, as already shown some downscaling methods do not work well in these catchments. See, for example, **page 17 lines 13 – 23 and page 21 lines 11 – 21.**

17. 7, 6188: “most likely” – check in the data

- The large EPI value found for this method in winter and in the Mulde catchment is due to the influence of two very large extremes created in the correction of the RCMs for the future. These are two events of 60 and 55 mm/day. From the BCM and BCQM the largest values obtained are approximately 40mm/day and the largest value obtained for the control period is 27mm/day. This is now described in a more clear way in the manuscript (**see page 20 lines 12 – 14**). In addition, the fact that BCMV may lead to unrealistic results is discussed in the selection of downscaling methods in the summary and conclusions section.

18. 21, 6188: “might be similar” – check in the data

- We have checked that there are very few rainy days in the catchment in Turkey as in the case of the catchment in Cyprus, only approximately 20% of days are rainy days. This has been added to the text, **see page 20 lines 29 – 31**.

19. 26-27, 6188: Why expect a larger impact in TR in summer?

- Because in this catchment and in summer there are only very few rainy days. This implies that in some cases all the rainy days in a season are included in the selection of extreme events. Therefore, in this catchment the change in the number of wet days may have an effect on the changes in extreme precipitation. CFQM and CFQP differ in the way that CFQP accounts for the change in the number of wet days. This discussion has been added to the manuscript, **see page 20 lines 29 - 31**.

20. 4.2: I think it would be more natural to have 4.2 before 4.1, i.e. first an evaluation in reference period and then results from future period.

- We prefer to keep 4.2 after 4.1. We consider that the use and comparison of the 8 statistical downscaling methods is the main part of this study. 4.2 considers only four of the methods, and is an additional analysis which is only possible for the bias correction methods.

21. 13, 6190: Not “extreme value index” but only “extreme values”, right?

- Corrected.

22. 5-15, 6191: Key paragraph which is far too compact. Only listing values is not very helpful; go much further, dig in the data, find out concretely why and under which circumstances the BC/DS methods decrease agreement with observations.

- We agree that this is an important point to highlight. In order to address it in more detail we have discussed the issue from the summary of the results for all the catchments (Figure 6) and then in detail for three of the catchments. The results obtained from the RCMs have

been added to Figure 7 in order to make this discussion possible. In general, in the catchments where the agreement between the observations and RCMs is good the downscaling methods used in this study are not able to improve it. In general, the simple BCM method tends to fail in improving the uncorrected RCMs more often than the other methods, but this should be tested for each case as it depends on the catchment. See discussion in the manuscript in **page 24 lines 2 – 12 page 25 lines 3 – 20**.

23. Table 3: Add the methods' abbreviations.

- The abbreviations of the downscaling methods have been added to Table 3.

24. Table 3: That seasonality is not taking into account in BCQM is an application issue, not a disadvantage of the method.

- This has been clarified in Table 3. In addition a note has been added to the table's caption to point out that some of the advantages and disadvantages are specific to the application, which is now indicated in all cases in Table 3.

25. Table 3: CFQP: ACF can be checked for all methods.

- This has been clarified in Table 3. See comment also reply to comment **24**.

26. Fig 5: Add legend, referring to other figure is not sufficient.

- The legend has now been added to both Figure 5 and 8.

Anonymous Referee #2

Received and published: 19 September 2014

The paper evaluates and compares 8 different statistical downscaling methods applied to 11 catchments in Europe using the ENSEMBLE projections. The study focused on extreme precipitation with return period of 1 and 5 years. The work is particularly useful for estimating hydrological impacts using climate model projections. I recommend moderate revision before it can be published.

General comments:

- 1.** The discussion or results lacks critical thinking and interpretation. It largely discusses the figures in text but does not always provide reasons. Sometimes, the authors are speculating, e.g. "This is likely due to the fact that in Europe, extreme precipitation from convective storms occurs more frequently during summer, and this has a larger influence on the outputs from the

RCMs and SDMs due to their higher spatial resolution.“ Do they know for sure extreme precipitation from convective storms occurs more frequently during summer in Europe and on what scales, daily, hourly? This needs to be supported by scientific evidence or citations.

- We agree that more interpretation of the results would improve the paper. We now discuss and provide interpretation when possible to a few key results (in agreement with the comments from reviewer #1). These are: the discussion on the difference between the climate change signal obtained from the downscaled series and the uncorrected RCMs (**this has been added see page 17 lines 13 – 23 and page 21 lines 11 – 21 and in agreement with the discussion in comment 16 from reviewer #1**); the fact that some bias correction methods do not improve the performance of the uncorrected RCMs (**this has been added in page 24 lines 2 – 12 page 25 lines 3 – 20 and in agreement with the discussion in comment 22 from reviewer #1**); and discussion of the similarities and differences of the downscaling methods (discussion added in the summary and conclusions section). Regarding the convective storms in summer, we have added a few references addressing daily extremes over Europe and the fact that RCMs tend to perform worse in summer due to the difficulties in representing convective storms:

Fowler, H. J. and Ekström, M.: Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes, *Int. J. Climatol.*, 29, 385–416, doi:10.1002/joc.1827, 2009.

Frei, C., Schöll, R., Fukutome, S., Schmidli, J. and Vidale, P. L.: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, *J. Geophys. Res.*, 111, D06105, doi:10.1029/2005JD005965, 2006.

Hofstra, N., Haylock, M., New, M. and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, *J. Geophys. Res.*, 114, D21101, doi:10.1029/2009JD011799, 2009.

Lenderink, G.: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations, *Clim. Res.*, 44, 151–166, doi:10.3354/cr00946, 2010.

2. The conclusion is not very useful after lots of work put in the comparison of 8 different statistical downscaling methods. The authors state in the conclusion: “we recommend the use of a set of statistical downscaling methods as well as an ensemble of climate model projections.”. But the authors also state: “There is not a tendency in the performance of the bias correction methods depending on the mean and extreme precipitation regime. It also shows that the differences between the methods and the performance of the bias correction methods depends on the catchment studied.” It does not provide any insight to people who want to conduct impact studies and need to choose downscaling methods for a river catchment. If the methods do not show much difference, why should we use a set of downscaling methods? And if we do have the research capacity to use several of them, what should we do with the set of different methods? E.g. Do we use the mean value? Are certain types of downscaling methods more suitable for certain catchments? It seems the authors have rushed to the conclusion without critical thinking and interpretation of the results.

- We have now extended the discussion on the selection of downscaling methods. It is difficult to point to a specific best method as it depends on several factors, but we have pointed out some issues that arise from the use of some methods (such as BCM, BCMV, CFM,...), which implies that they should not be selected for some specific applications. For example, in the case of BCMV it might lead to unrealistic precipitation events; CFM does not account for the fact that the change in extreme precipitation might be different than the change in mean precipitation; and BCM is in many cases not able to improve the representation of extremes compared to the uncorrected RCMs. Even if methods do not show large differences, if they are based on different assumptions they should be included in the ensemble of methods. When possible, we recommend using an ensemble of methods to be able to assess the uncertainty in extreme precipitation projections. **See summary of this discussion in page 27 line 13 – page 28 line 4 and page 28 line 11 – 15 and in the different sections in the manuscript in agreement with comments from both reviewer #1 and #2.**

3. The paper is full of acronyms (names of catchments and methods) which make it very hard to follow.

- We agree that there are many acronyms in the paper and that in some cases it might be difficult to follow for the reader. In the revised version of the manuscript, the acronyms for the catchments have finally been maintained (see **“General comments” in this document**). However, the acronyms for the statistical downscaling methods have been kept.

Specific comments:

4. 6169-23: ‘up to half’ is not strictly correct according to Figure 3. In the conclusion, authors use ‘at least 30%’ instead of ‘up to half’. The former is better.

- The variance explained by the SDMs ranges from 30% to slightly more than 50%. It has now been made clearer in the text (see **page 2 line 19**) that at least 30% and up to approximately 50% of the total variance is due to the statistical downscaling methods used.

5. 6170-18: ‘approximately 25km’ but EURO-CORDEX is at a finer resolution.

- True. It has been clarified in the text that the new EURO-CORDEX simulations have a higher spatial resolution (11 km) than most of the RCMs available which often have a resolution of approximately 25 km. **See page 3 line 16 – 17.**

6. 6171-15: ‘Wetterhall et al. (2012) Conditioning model output statistics of regional climate model precipitation on circulation patterns’ provides another interesting comparison, worth adding to the literature review.

- Wetterhall et al. (2012) provides an interesting comparison of downscaling methods. Their results showing that it was not possible to reject the hypothesis that all the SDMs perform equally well is interesting and relevant for this paper. This study has been included in the literature review in **page 4 line 32**.

7. 6172-8: Hundedcha et al., 2014. Is this already accepted or published?

- The paper was submitted in September 2014.

8. 6172-24: ‘where most extremes occur’. Throughout the paper, authors have not defined the extreme precipitation. Is it based on return period or percentile?

- The season with more occurrences of extremes is estimated using the Extreme Precipitation Index. First the extreme events for the observational data sets are obtained using the 1 year return period threshold for the whole period. Then the season with more extremes is estimated from this extreme value series. Throughout the paper extremes are defined as described in the section “Extreme Precipitation Index”. This has been clarified in the text in **page 6 line 13 – 15**.

9. 6173-5: ‘while the other catchments use’, change to ‘while the remaining use’

- Corrected for “while the remaining ones use” (**page 6 line 21**).

10. 6176-11: ‘closed from’ change to ‘closed form’

- Corrected.

11. 6181-16/24: (i) and (iii) should be (1) and (3)

- Corrected.

12. 6184-3: ‘three source’ should be ‘three sources’

- Corrected.

13. 6193-9: ‘for higher return levels’. There are only 1 and 5 years return levels presented in this study, so this should be ‘for the higher return level, 5 years’.

- Corrected.

1 **Inter-comparison of statistical downscaling methods for**
2 **projection of extreme precipitation in Europe**

3

4 **M. A. Sunyer¹, Y. Hundecha², D. Lawrence³, H. Madsen⁴, P. Willems^{5,6}, M.**
5 **Martinkova⁷, K. Vormoor⁸, G. Bürger⁸, M. Hanel^{7,9}, J. Kriaučiūnienė¹⁰, A. Loukas¹¹, M.**
6 **Osuch¹², and I. Yücel¹³**

7

8 [1]{DTU Environment, Lyngby, Denmark}

9 [2]{Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum, Potsdam, Germany}

10 [3]{Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway}

11 [4]{DHI, Hørsholm, Denmark}

12 [5]{Hydraulics Laboratory, KU Leuven, Heverlee, Belgium}

13 [6]{Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels,
14 Belgium}

15 [7]{Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech
16 Republic}

17 [8]{Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany}

18 [9]{T. G. Masaryk Water Research Institute, Prague, Czech Republic}

19 [10]{Lithuanian Energy Institute, Kaunas, Lithuania}

20 [11]{Department of Civil Engineering, University of Thessaly, Volos, Greece}

21 [12]{Department of Hydrology and Hydrodynamics, Institute of Geophysics, Polish Academy of
22 Sciences, Warsaw, Poland}

23 [13]{Middle East Technical University, Civil Engineering Department, Ankara, Turkey}

24

25 Correspondence to: M.A. Sunyer (masu@env.dtu.dk)

1 **Abstract**

2 Information on extreme precipitation for future climate is needed to assess the changes in the
3 frequency and intensity of flooding. The primary source of information in climate change impact
4 studies is climate model projections. However, due to the coarse resolution and biases of these
5 models, they cannot be directly used in hydrological models. Hence, statistical downscaling is
6 necessary to address climate change impacts at the catchment scale.

7 This study compares eight statistical downscaling methods often used in climate change impact
8 studies. Four methods are based on change factors, three are bias correction methods, and one is a
9 perfect prognosis method. The eight methods are used to downscale precipitation output from
10 fifteen regional climate models (RCMs) from the ENSEMBLES project for eleven catchments in
11 Europe. The overall results point to an increase in extreme precipitation in most catchments in both
12 winter and summer. For individual catchments, the downscaled time series tend to agree on the
13 direction of the change but differ in the magnitude. Differences between the statistical downscaling
14 methods vary between the catchments and depend on the season analysed. Similarly, general
15 conclusions cannot be drawn regarding the differences between change factor and bias correction
16 methods. The performance of the bias correction methods during the control period also depends on
17 the catchment, but in most cases they represent an improvement compared to RCM outputs.
18 Analysis of the variance in the ensemble of RCMs and statistical downscaling methods indicates
19 that at least 30% and up to approximately half of the total variance is derived from the statistical
20 downscaling methods. This study illustrates the large variability in the expected changes in extreme
21 precipitation and highlights the need of considering an ensemble of both statistical downscaling
22 methods and climate models. Recommendations are provided on selection of the most suitable
23 statistical downscaling methods to include in the analysis.

24

1. Introduction

Both the frequency and intensity of extreme precipitation are expected to increase under climate change conditions in Europe (Christensen and Christensen, 2003; IPCC, 2012). Several climate studies have focused on assessing these changes (e.g. Fowler and Ekström, 2009; Frei et al., 2006; Kendon et al., 2008) and their consequences in relation to the risk of flooding (Christensen and Christensen, 2003; IPCC, 2012; Leander et al., 2008; Vansteenkiste et al., 2013). The main steps often followed in these studies comprise the selection of one or several global climate models (GCM), regional climate models (RCM) and/or statistical downscaling methods (SDM). In climate change impact studies, hydrological models are then used to estimate changes in hydrological variables.

GCMs are the most comprehensive and widely used models for simulating the response of the global climate system to changes in greenhouse gas emissions. However, their spatial resolution (approximately 150 km) is often too coarse for addressing climate change impacts at the local scale, and variables such as precipitation are often biased. RCMs are climate models that cover a specific region (e.g. Europe) and use GCMs as boundary condition. RCMs have a higher spatial resolution (often approximately 25 km, but the new EURO-CORDEX simulations (Jacob et al., 2013) have a resolution of approximately 11 km) than GCMs, which makes them more adequate for assessing changes at the local scale. Nonetheless, RCMs often inherit the biases from the GCMs and their spatial resolution might still be too coarse for some impact studies (Maraun et al., 2010). Hence, further statistical downscaling is often needed to obtain bias-corrected projections at the local scale (Fowler et al., 2007). Statistical downscaling is based on defining a relationship between the large scale outputs of the RCMs (or GCMs) and the local scale variables required in impact studies (Fowler et al., 2007; Wilby et al., 2004).

In recent years, a relatively large number of RCM outputs have been made available, but there is no consensus on the best way to assess their performance (Knutti et al., 2010). There are several challenges in evaluating RCMs. For example, a RCM might perform well for some variables in some regions but not for other variables. Moreover, even if a climate model performs well under present climate conditions it might not perform equally well under future conditions (Knutti, 2010). For these reasons, it is generally recommended to use a multi-model ensemble of RCMs (or GCMs) instead of using a single model (Knutti et al., 2010; van der Linden and Mitchell, 2009; Tebaldi and Knutti, 2007).

1 Similarly, a large number of SDMs have been suggested in the literature, but there is no consensus
2 on the best SDM. Fowler et al. (2007) and Maraun et al. (2010) provide comprehensive reviews of
3 the methods suggested in the literature and their suitability for different applications. As in the case
4 of climate models, the validation of SDMs is challenging. Only a few recent studies address this
5 issue (e.g. Maraun et al., 2013; Räisänen and Rätty, 2013; Teutschbein and Seibert, 2012; Vrac et
6 al., 2007).

7 In order to account for the uncertainties in climate change impact studies and due to the lack of
8 consensus on the best climate model and SDM, a number of studies consider multiple climate
9 models and SDMs. For example regarding extreme events, Bürger et al. (2012 and 2013) used eight
10 SDMs to downscale six GCMs forced with three emission scenarios, Sunyer et al. (2012) used five
11 SDMs to downscale four RCMs driven by two GCMs, Hanel et al. (2013) used four SDMs and
12 fifteen RCMs, and Kidmose et al. (2013) used two SDMs and nine RCMs. Bürger et al. (2012 and
13 2013) assessed the performance and variance arising from the SDMs and GCMs. They concluded
14 that the main influence on the overall results for different extreme indices (including both
15 precipitation and temperature indices) was the downscaling method used followed by the climate
16 model selected. In their study, the main source of variance depended on the index considered, but
17 overall the climate models had more influence on precipitation than on temperature indices. Sunyer
18 et al. (2012) and Hanel et al. (2013) showed that the variation in the results arising from the use of
19 several statistical downscaling methods is larger in the case of extreme events (extreme
20 precipitation in the case of Sunyer et al. (2012) and droughts in the case of Hanel et al. (2013)).
21 Kidmose et al. (2013) found that in the case of extreme groundwater levels in Denmark the variance
22 arising from the RCMs was larger than from the SDMs, but in this case only two SDMs were
23 considered.

24 Some studies also consider hydrological models in the chain of uncertainties. For example, Wilby
25 and Harris (2006) used two SDMs, four GCMs, and two emission scenarios combined with two
26 hydrological model structures and two sets of hydrological model parameters. They concluded that
27 the main sources of variation in the case of low flows are associated with the SDMs and GCMs
28 used. Lawrence and Haddeland (2011) compared two SDMs, six RCMs driven by two GCMs, and
29 two emission scenarios and used multiple parameter sets for the hydrological impact model. They
30 found that for rainfall dominated catchments, the uncertainty arising from the hydrological
31 parameters was more significant than other sources. In snowmelt dominated catchments, however,
32 climate scenarios and SDMs were the main source of uncertainty. Wetterhall et al. (2012) assessed

1 the variability in extreme discharge using three SDMs, sixteen RCMs, one hydrological model and
2 a set of model parameters. The performance of the SDMs was evaluated and a best method was
3 found, but it was not possible to reject the hypothesis that all SDMs perform equally well.
4 Wetterhall et al. (2012) also concluded that more complex SDMs performed better than simple
5 methods. A similar conclusion was reached by Rätty et al. (2014) and Teutschbein and Seibert
6 (2012). These two studies mainly focused on the validation of SDMs. Teutschbein and Seibert
7 (2012) considered six SDMs and eleven RCMs for five Swedish catchments, while Rätty et al.
8 (2014) considered nine SDMs and six RCMs and considered two regions, Northern and Southern
9 Europe.

10 The main focus of this study is to assess and compare the changes in extreme precipitation obtained
11 using a range of SDMs and RCMs in eleven European catchments. For this purpose, precipitation
12 outputs from fifteen RCMs driven by six GCMs from the ENSEMBLES project (van der Linden
13 and Mitchell, 2009) are downscaled using eight SDMs based on different underlying assumptions.
14 Four SDMs are change factor methods, three are bias correction methods and one is a perfect
15 prognosis method. Some previous studies have compared the results from change factors and bias
16 correction methods (e.g. Hanel et al., 2013; Ho et al., 2012; Räisänen and Rätty, 2013) for mean
17 temperature and mean precipitation for specific catchments. Here we focus on changes in extreme
18 precipitation in a range of catchments over Europe with different climates. A key objective of this
19 study is to assess whether it is possible to identify general advantages and deficiencies of the
20 different SDMs when applied to the different catchments, and hence outline recommended use of
21 SDMs. In addition, this study also focuses on whether there are common trends in projected
22 changes in extreme precipitation over Europe and what the main sources of variation in the changes
23 in extreme precipitation are.

24 The results presented here are based on a coordinated effort carried out as part of the COST Action
25 FloodFreq (European Procedures for Flood Frequency Estimation, www.cost-floodfreq.eu). The
26 outputs from this study have also been used as inputs to hydrological impact modelling in order to
27 assess the changes in extreme discharge and flood frequency in the eleven catchments (Hundecha et
28 al., submitted).

29 The next section describes the case study catchments and the data used, followed by the
30 methodology section. Section 4 presents and discusses the results, and section 5 summarizes the
31 findings and conclusions of the study.

2. Case study catchments and data

2.1. Observations

Figure 1 shows the location of the eleven catchments studied and the main properties of each catchment are summarized in Table 1. The two most northern catchments are the Norwegian catchments Nordelva at Krinsvatn (NO2) and Atna at Atnasjø (NO1), and the most southern catchment is Yermasoyia (CY) in Cyprus. The size of the catchments varies from the 6171 km² of Mulde (DE) in Germany to the 67 km² of Upper Metuje (CZ2) in the Czech Republic. Different precipitation patterns are represented in the catchments. The mean precipitation ranges between 2437 mm yr⁻¹ in NO2 to 589 mm yr⁻¹ in Nysa Kłodzka in Poland (PL). The season with more extreme precipitation events is summer for most of the catchments: NO1, DE, Aarhus in Denmark (DK), Merkys in Lithuania (LT), Grote Nete in Belgium (BE), and Jizera in the Czech Republic (CZ1). In NO2 and CY, winter is the season where most extremes occur, while in the Turkish catchment Omerli (TR) it is autumn. The season which is most subject to extremes is estimated from the extreme value series obtained considering the 1-yr threshold level and the whole time series (see section 3.2 for more details on how extreme precipitation is defined).

FIGURE 1

The observational data used is daily catchment precipitation, as the data were to be further used in catchment-based hydrological modelling in separate work (Hundecha et al., submitted). Different methods have been used to obtain areal precipitation time series. The catchments NO2, NO1, DK, and CZ2 use gridded data (derived from station data) to obtain areal average daily values for the catchment, while the remaining ones use station data to construct areal values. The cut-off value (threshold for dry days) for the observational data differs somewhat between the catchments. These catchment specific thresholds were not applied to the RCMs as they are not considered relevant for the analysis of extreme precipitation. Nonetheless some of the SDMs use thresholds to define dry and wet days (see section 3).

TABLE 1

2.2. Regional Climate Models

The climate model data used in this study is an ensemble of fifteen RCMs from the ENSEMBLES project (van der Linden and Mitchell, 2009). These fifteen simulations are based on eleven RCMs driven by six different GCMs. Table 2 shows the combinations of RCMs-GCMs used. The spatial

1 resolution of all the models is 0.22° (approximately 25 km). For all the models, daily precipitation
2 time series are available for the time period 1951–2100. In this study, we consider the time period
3 1961–1990 and 2071–2100 as the control and future time periods, respectively. It must be noted that
4 six RCMs do not have data available for the year 2100. The future period used for these models is
5 2071–2099; this is not expected to have an influence on the results of this study. For each
6 catchment, daily precipitation has been extracted from the 15 RCMs for the two periods using
7 nearest neighbour interpolation to the catchment centroid. It must be noted that to simplify the
8 calculations, the same control period is used for all the catchments. Therefore, in some catchments,
9 the time period with observations (see Table 1) and the control period used from the RCMs do not
10 fully overlap.

11 TABLE 2

12 3. Methodology

13 3.1. Statistical downscaling methods

14 Eight SDMs are used to obtain downscaled RCM projections at the catchment scale. These methods
15 are based on the idea that it is possible to define a relationship between the large scale variables
16 (RCM outputs) and local scale variables (catchment precipitation). Wilby and Wigely (1997) and
17 Fowler et al. (2007) classify SDMs based on the relationship used to link large and local scale. They
18 consider three groups: regression methods, weather type approaches and stochastic weather
19 generators. Rummukainen (1997) classifies SDMs based on the information used from the large
20 scale variables and defines two groups: perfect prognosis (PP) and model output statistics (MOS).
21 Maraun et al. (2010) integrate both Rummukainen (1997) and Wilby and Wigely (1997)
22 classifications and consider three groups: PP, MOS, and weather generators. According to this last
23 classification, seven of the eight methods used here are MOS methods, and one method is a PP
24 method.

25 Here we further classify the seven MOS methods into change factor (CF) methods and bias
26 correction (BC) methods. Four of the MOS methods considered are CF and three are BC methods.
27 CF methods estimate the change from control to future period projected by the RCM in one or
28 several statistics and apply this change to the observations. These methods are based on the idea
29 that RCMs represent the change from the control to the future climate better than the absolute
30 values of the variables. The BC methods define a transfer function for the RCM outputs for the
31 control period to match certain statistical properties of the observations. This transfer function is

1 then used to correct the RCM outputs for the future period. CF methods preserve the temporal
2 structure in the observed time series while BC methods preserve the temporal structure in the RCM
3 outputs. It must be noted that both approaches are based on the assumption that the bias for the
4 future period is identical to the bias for the control period, which may not be the case. Sunyer et al.
5 (2014) show that the precipitation bias of the RCMs depends on the precipitation intensity and
6 might change in the future.

7 The following sub-sections briefly describe the eight SDMs. In the results section we refer to the
8 SDMs as either CF or BC methods. For simplicity, the perfect prognosis method is grouped with
9 the BC methods even though it does not strictly correct the RCMs. It is included with the BC
10 methods because it defines a transfer function between the RCM for the control period and the
11 observations and then applies this to the RCM output for the future period.

12 A common terminology is used for describing the methods: P^{Obs} and P^{Fut} refer to the observed
13 precipitation and the downscaled precipitation for the future period, respectively; and P^{RCMCon} and
14 P^{RCMFut} refer to the precipitation output from the RCMs for the control and future time period,
15 respectively. Similarly, ECDF^{Obs} and ECDF^{Fut} refer to the empirical cumulative distribution
16 function (ECDF) for the observed precipitation and for the downscaled precipitation for the future
17 while $\text{ECDF}^{\text{RCMCon}}$ and $\text{ECDF}^{\text{RCMFut}}$ refer to the ECDF estimated from the RCMs for control and
18 future time period, respectively. The methods used here have been implemented as suggested in the
19 literature, i.e. no harmonisation has been applied to enable, for example, a common method for
20 accounting for seasonality or the definition of wet days. This is due to this study focused on the
21 intercomparison of approaches in the way they are applied by the partners of FloodFreq COST
22 Action, which was designed for the exchange and compilation of ideas and knowledge across
23 participating countries. Table 3 summarizes the main advantages and disadvantages of each method.

24 3.1.1. Bias correction of mean

25 The bias correction of mean, BCM, is a simple method based on removing systematic errors in
26 mean daily precipitation. It has been used in several hydrological applications (e.g. Hanel et al.,
27 2013; Leander and Buishand, 2007; Leander et al., 2008). Here the method proposed by Leander
28 and Buishand (2007) is used. This is based on the transformation:

$$29 \quad P_{y,j}^{\text{Fut}} = a_j P_{y,j}^{\text{RCMFut}} \quad (1)$$

1 where y is the year, j is the day of the year and a_j is the transformation parameter. a_j is estimated in
 2 two steps. First, for all the years a subset of 61 days centred on day j is created for $P_{.,j}^{\text{Obs}}$ and
 3 $P_{.,j}^{\text{RCMCon}}$. Then, a_j is estimated as the mean of $P_{.,j}^{\text{Obs}}$ divided by the mean of $P_{.,j}^{\text{RCMCon}}$.

4 **3.1.2. Bias correction of mean and variance**

5 The bias correction of mean and variance method, BCMV, is an extension of the BCM method. It
 6 corrects the RCM outputs considering systematic errors in both the mean and the variance. This
 7 method has been applied in several studies (e.g. Hanel et al., 2013; Leander and Buishand, 2007;
 8 Leander et al., 2008). The method suggested by Leander and Buishand (2007) is followed here,
 9 which is based on the transformation:

$$10 \quad P_{y,j}^{\text{Fut}} = a_j \left(P_{y,j}^{\text{RCMFut}} \right)^{b_j} \quad (2)$$

11 where a_j is estimated as described above for BCM, and b_j is estimated by equating the coefficient
 12 of variation of $(a_j P_{.,j}^{\text{RCMCon}})^{b_j}$ and $P_{.,j}^{\text{Obs}}$. b_j is found by iteration since it is not possible to solve this
 13 equation in closed **form**.

14 **3.1.3. Bias correction quantile mapping**

15 Bias correction based on quantile mapping, BCQM, has been widely used to correct RCM outputs
 16 over Europe (e.g. Dosio and Paruolo, 2011; Gudmundsson et al., 2012; Piani et al., 2010). The non-
 17 parametric empirical quantile method suggested in Gudmundsson et al. (2012) is followed here. It is
 18 based on the concept that there exists a transformation h , such that:

$$19 \quad P^{\text{Obs}} = h \left(P^{\text{RCMCon}} \right) = ECDF^{\text{Obs}}^{-1} \left(ECDF^{\text{RCMCon}} \left(P^{\text{RCMCon}} \right) \right) \quad (3)$$

20 First, all the probabilities in $ECDF^{\text{Obs}}$ and $ECDF^{\text{RCMCon}}$ are estimated at a fixed interval of 0.01.
 21 Then, h is estimated as the relative difference between the two ECDFs in each interval.
 22 Interpolation between the fixed intervals is based on a monotonic tricubic spline interpolation. A
 23 threshold for the correction of the number of wet days is estimated from the empirical probability of
 24 non-zero values in P^{Obs} . All RCM values below this threshold are set to zero. The precipitation
 25 values for the full annual daily series are corrected without subsampling by season or month, as
 26 suggested by Piani et al., 2010. The method was implemented in R using the *qmap* package
 27 (Gudmundsson, 2014).

28 **3.1.4. Expanded downscaling**

1 Expanded Downscaling, XDS, is a perfect prognosis technique which maps large-scale atmospheric
2 fields to local station data. XDS was originally introduced for weather forecasting purposes, but it
3 has been recently used in climate change studies (e.g. Bürger and Chen, 2005; Bürger et al., 2013;
4 Dobler et al., 2012). The XDS approach is based on defining a multivariate linear regression
5 between predictors y (multivariate fields of atmospheric variables) and predictands x (local scale
6 variables, i.e. catchment precipitation), extended by the side condition that the local co-variability
7 between the variables (and stations) is preserved:

$$8 \quad XDS = \arg \min_Q \|xQ - y\|, \text{ subject to } Q'x'xQ = y'y, \quad (4)$$

9 where XDS is the least square-solution of the matrix Q which is found among those that preserve
10 the local covariance ($Q'x'xQ = y'y$). By this approach, the estimation of extremes is supposed to
11 be improved compared to regular linear regression models. See Bürger et al. (2009) for a detailed
12 description of this method.

13 The XDS model is first trained on RCM atmospheric fields driven by the ECMWF ERA-40
14 reanalysis (Uppala et al., 2005) and local scale observations with at least 10 yrs of data. Then, RCM
15 outputs for the control and future periods are used to generate time series at the local scale.
16 Generally XDS allows for exploring a range of large scale variables as predictors. Large-scale
17 reanalyses, however, are generally in better agreement with local observations than an RCM
18 simulation driven by those reanalyses, simply because that the simulation likely differs from the
19 actual weather realization which is used for XDS calibration. This has the consequence that a
20 perfect prognosis approach is no longer perfect. A second data assimilation based on the RCM-
21 ERA-40 runs (in addition to the data assimilation which has already been done for the ERA-40
22 reanalysis) would overcome this problem to some degree. However, such runs are not available for
23 the RCMs accessible from the ENSEMBLES archive. For this study, the predictors were therefore
24 chosen rather 'conservatively', with predictor variables being limited to large-scale total and
25 convective precipitation. The result is a set of predictors that is, moreover, unique across all
26 catchments. The XDS source code and documentation can be downloaded from:
27 <http://xds.googlecode.com>.

28 **3.1.5. Change factor of mean**

29 The change factor of mean, CFM, is a simple method which has been widely applied in
30 hydrological applications (Hanel et al., 2013; Prudhomme et al., 2002; Sunyer et al., 2012). It is

1 based on applying the change in mean precipitation projected by the RCMs to the observed data.
 2 The method described in Sunyer et al. (2012) is followed here. Similarly to BCM, this method is
 3 based on the transformation:

$$4 \quad P_{m,t}^{\text{Fut}} = a_m P_{m,t}^{\text{Obs}} \quad (5)$$

5 where m refers to the month and t to each time step in the observations; a_m is the relative change in
 6 the precipitation mean for month m . a_m is estimated as the mean of $P_{m,\cdot}^{\text{RCMFut}}$ divided by the mean
 7 of $P_{m,\cdot}^{\text{RCMCon}}$.

8 **3.1.6. Change factor of mean and variance**

9 The change factor of mean and variance, CFMV, is an extension of CFM. It has been applied in
 10 several studies (e.g. Hanel et al., 2013; Räisänen and Rätty, 2013; Sunyer et al., 2012). CFMV
 11 modifies the observed time series using the change in both the mean and variance. The method
 12 described in Sunyer et al. (2012) is followed here. Similar to BCMV, the method is based on the
 13 transformation:

$$14 \quad P_{m,t}^{\text{Fut}} = a_m \left(P_{m,t}^{\text{Obs}} \right)^{b_m} \quad (6)$$

15 where a_m is estimated as described for CFM. b_m is estimated by equating the coefficient of variation
 16 of the time series $(a_m P_{m,\cdot}^{\text{Obs}})^{b_m}$ and the coefficient of variation estimated for the future period. As in
 17 BCMV, this is solved by iteration. The coefficient of variation for the future period is calculated
 18 from the relative change in the mean and variance projected by the RCMs.

19 **3.1.7. Change factor quantile mapping**

20 The change factor quantile mapping, CFQM, is based on using the relative change in the ECDF
 21 projected by the RCMs to modify the observed data. It has been applied in several climate change
 22 studies (e.g. Boé et al., 2007; Olsson et al., 2009).

23 This method uses the ECDF of wet days estimated for each month m for the observations, and the
 24 RCM output for the control and future periods. The probability intervals considered are 0.001 for
 25 quantiles lower than 0.9 and 0.0005 for higher quantiles (linear interpolation between intensities is
 26 applied to obtain the precipitation intensity for all the quantiles). Wet days are defined as days with
 27 precipitation higher than 1 mm. The perturbation of the observed time series is carried out in three
 28 steps. First, for each wet day in each month m , $\text{ECDF}_m^{\text{Obs}}$ is used to estimate the probability of the
 29 precipitation intensity. Second, the relative change in the intensity for this probability is estimated

1 from $ECDF_m^{RCMFut}$ and $ECDF_m^{RCMCon}$. This change is then multiplied to the observed precipitation
2 intensity to obtain the intensity for the future period. Dry days in the observations are not modified.

3 **3.1.8. Change factor quantile perturbation**

4 The change factor quantile perturbation, CFQP, is similar to CFQM but it also accounts for changes
5 in the number of wet days. Quantile perturbation methods can be performed either in a non-
6 parametric way (Ntegeka et al., 2014; Vansteenkiste et al., 2014; Taye et al., 2011; Willems and
7 Vrac, 2011) or in a parametric way based on distribution calibration (Willems, 2013; Rana et al.,
8 2014). The version used here is the non-parametric one that was applied by Willems and Vrac
9 (2011).

10 The observations are perturbed using a two-step approach. First, the number of wet days (days with
11 precipitation higher than 0.1 mm day^{-1}) is changed for each month. The relative change in the
12 frequency of wet days is estimated from the RCM output. If the frequency increases, dry days are
13 randomly selected and replaced by random wet day intensities from the time series. Otherwise, wet
14 days are randomly replaced by zero precipitation. In the second step, the wet day intensities are
15 perturbed in a similar way as in the CFQM method. The empirical probability of each intensity is
16 estimated, and the relative change in the intensity for each probability is then calculated (linear
17 interpolation is applied when different probabilities are obtained for the control and future period)
18 and used to perturb the observations.

19 These two steps are repeated 10 times. The repetition that leads to the results closest to the mean
20 monthly precipitation value of all the repetitions is selected. See Willems and Vrac (2011) for more
21 details on this method, including checks of the coefficient of variation, skewness and
22 autocorrelation for the results.

23 It must be noted that in the case of BCQM, CFQM, and CFQP the use of empirical quantiles may
24 lead to large fluctuations representing a lack of robustness in the values of the CF (or correction
25 factors in the case of BCQM) for the highest quantiles. This is due to the fact that the highest
26 quantiles are estimated using a limited number of values.

27 **3.2. Extreme precipitation Index**

28 The outputs from all the statistical downscaling methods are analysed using an extreme
29 precipitation index (EPI). This is defined as the average change in extreme precipitation higher than
30 a defined return period. In this study, the return period is set equal to 1 and 5 yrs. EPI is estimated

1 separately for each SDM, RCM, catchment, threshold return period, season and temporal
2 aggregation. Four seasons are considered: winter (December to February), spring (March to May),
3 summer (June to August), and autumn (September to November). Additionally, the index is
4 estimated considering the whole time series, i.e. without dividing in seasons. The temporal
5 aggregations considered are 1, 2, 5, 10, and 30 days. These are estimated using a moving average
6 from the daily time series.

7 The first step in the calculation of EPI is the extraction of the extreme value series from the
8 precipitation time series using a Peak Over Threshold (POT) approach. Peaks are extracted by using
9 the 1- and 5-yr threshold return periods. For example, with a 30-yr record, the 30 and 6 most
10 extreme events are included in the extreme series for the 1- and 5-yr threshold levels, respectively.
11 An independence criterion based on the inter-event time is applied to make sure that extreme values
12 are independent, i.e. only values separated by more than Δt days are considered. Δt is set equal to
13 the temporal aggregation, i.e. for an aggregation time of 1 day, events must be separated by more
14 than one day. EPI is then estimated as:

$$15 \quad \text{EPI} = \frac{\overline{\text{POT}}_2}{\overline{\text{POT}}_1} \quad (7)$$

16 where $\overline{\text{POT}}_1$ and $\overline{\text{POT}}_2$ are the averages of the selected POT values for reference and scenario,
17 respectively. EPI takes the value of 1 if no change is estimated from reference to scenario and
18 greater (less) than 1 if the average extreme precipitation is higher (lower) in the scenario time
19 series.

20 In the results section, EPI is used to compare the changes in the downscaled time series from
21 control to future. Additionally, three further comparisons are carried out. In total EPI is calculated
22 for four different cases:

- 23 1. Comparison of the downscaled time series for the control and future periods.
- 24 2. Comparison of the RCM outputs for control and future periods. This allows us to compare
25 the changes estimated from the downscaled precipitation, estimated in (1), to the changes
26 projected by the RCMs.
- 27 3. For the four BC methods: comparison of the observations and the bias corrected RCMs for
28 the control period. The value of the index for this comparison is a measure of the error of the
29 BC methods in bias correcting the RCM outputs for extreme precipitation.

4. Comparison of the observations and RCM outputs for the control period. This comparison evaluates the performance of the RCMs in simulating extreme precipitation, and allows us to assess whether the error in the bias corrected time series, estimated in (3), is smaller than in the RCMs.

3.3. Variance decomposition

The variability in the EPI values found when comparing the downscaled time series for control and future arises mainly from three sources: GCMs, RCMs and SDMs. A variance decomposition approach is used to address the influence of each of these sources on the total variance for each catchment, return level, season and temporal aggregation. The approach described in Déqué et al. (2007, 2012) is followed here.

The total variance of EPI, V , can be split into the different contributions as:

$$V = R + G + S + RG + RS + GS + RGS \quad (8)$$

where R , G , and S are the individual parts of the variance explained by the RCMs, GCMs, and SDMs, respectively; RG , RS , and GS are the variance due to the interaction of RCM-GCM, RCM-SDM, and GCM-SDM, respectively; and RGS is the variance due to the interaction of all three sources. The part of the total variance explained by the RCMs, $V(R)$ is:

$$V(R) = R + RG + RS + RGS \quad (9)$$

The part of the total variance due to the GCMs, $V(G)$, and SDMs, $V(S)$, can be obtained in a similar way. The variances in Eq. (8) and Eq. (9) can be estimated as:

$$\begin{aligned} R &= \frac{1}{11} \sum_{i=1}^{11} (\overline{\text{EPI}}_{i..} - \overline{\text{EPI}}_{...})^2; \\ RG &= \frac{1}{11} \frac{1}{6} \sum_{i=1}^{11} \sum_{j=1}^6 (\overline{\text{EPI}}_{ij.} - \overline{\text{EPI}}_{i..} - \overline{\text{EPI}}_{.j.} + \overline{\text{EPI}}_{...})^2 \\ RGS &= \frac{1}{11} \frac{1}{6} \frac{1}{8} \sum_{i=1}^{11} \sum_{j=1}^6 \sum_{k=1}^8 (\overline{\text{EPI}}_{ijk} - \overline{\text{EPI}}_{ij.} - \overline{\text{EPI}}_{i.k} - \overline{\text{EPI}}_{.jk} + \overline{\text{EPI}}_{i..} + \overline{\text{EPI}}_{.j.} + \overline{\text{EPI}}_{.k.} - \overline{\text{EPI}}_{...})^2 \end{aligned} \quad (10)$$

where EPI_{ijk} is value of the index for RCM i , GCM j and SDM k , $\overline{\text{EPI}}$ represents the average of EPI with respect to the subscripts that are replaced by a dot. The rest of the terms in Eq. (9) are estimated in a similar way as shown in Eq. (10). For more details see Déqué et al. (2007, 2012). Note that the observation errors in this approach are neglected in comparison with the other error sources.

1 As in Déqué et al. (2007), not all the terms in Eq. (10) can be estimated. This is because not all the
2 combinations of RCM-GCMs are available (see Table 2). Déqué et al. (2007) suggested a simple
3 method to reconstruct the missing data in the matrix of RCM-GCMs. This is based on minimizing
4 the full interaction term RGS . However, this approach cannot be directly used here. This is because
5 for the combinations of RCM i and GCM j that are not available there is no information on any of
6 these SDM k values. Hence, in some cases it is not possible to estimate \overline{EPI}_{ij} , which is needed to
7 minimize the full interaction term RGS . For this reason, a slight modification is made to the
8 approach suggested by Déqué et al. (2007). The approach followed here consists of two steps: (i)
9 for all the combinations of i and j missing, \overline{EPI}_{ij} is estimated by minimizing RG ; and (ii) the values
10 of EPI_{ijk} missing are estimated by minimizing RGS .

11 A large number of gaps must be filled using this procedure. Two simple verifications have been
12 carried out to check that the results are not largely affected by the matrix reconstruction approach.
13 The first verification procedure is a simple comparison of the results from the variance
14 decomposition described above with a variance decomposition approach, which considers only two
15 sources of variance (climate models and SDMs). In the approach considering only these two
16 sources, matrix reconstruction is not needed because all the elements in the matrix are known. The
17 second verification procedure is similar to the verification carried out in Déqué et al. (2007). The
18 two verification approaches and their results are described in Appendix A.

19 The results from the first verification procedure show that the conclusion as to which is the most
20 important source of variance is nearly the same when considering two or three sources for all
21 catchments. Conversely, the results from the second verification show that the reconstruction
22 approach can influence the results. From the results of the first verification, we decide to analyse the
23 variance explained by the GCMs and RCMs separately (i.e. considering three sources of variance)
24 because, in our opinion, it adds value to separate the influence of the GCMs and RCMs.
25 Nonetheless, we acknowledge that the results must be treated with caution due to the uncertainty
26 added in the matrix reconstruction procedure.

27

28

TABLE 3

1 **4. Results and discussion**

2 This section is divided into two main parts. The first part analyses the results of all SDMs. The
3 second part focuses on the performance of the three BC methods and perfect prognosis method. All
4 the results are shown for winter and summer as these are the two seasons where most of the
5 extremes occur under present conditions. However, it should be noted that in some catchments
6 changes in other seasons might also be important due to their influence on floods, see examples in
7 Hundedcha et al. (submitted).

8 **4.1. Comparison of the downscaled time series for the control and future periods**

9 This subsection analyses the results of the eight SDMs driven by all RCMs. A summary of the
10 results obtained for all the catchments is first presented followed by a more detailed analysis of the
11 differences between the SDMs for three selected catchments.

12 **4.1.1. Extreme precipitation index and variance decomposition for all catchments**

13 Figure 2 summarizes the results of all the SDMs and RCMs for all the catchments for winter and
14 summer for a temporal aggregation of 1 day. Additionally, it compares the results of the SDMs with
15 the changes between the control and future periods projected by the RCMs. For the catchment CY
16 for some SDMs, two special situations are encountered. For the methods BCM and BCMV for both
17 winter and summer periods, due to the few rainy days in some of the RCM simulations, some of the
18 parameters take unrealistic values which lead to unrealistic values of EPI. Similarly, it is not
19 possible to estimate the CFs used in the case of CFM, CFMV and CFQM in the summer period.
20 The results of these methods are, therefore, not included in the analysis for CY. For the other
21 catchments such problems with the SDMs were not encountered and all results are included in the
22 analysis.

23 For winter, extreme precipitation is expected to increase in all catchments (the median of EPI is
24 greater than 1) except in CY. The median of EPI is similar for all catchments except for the two
25 most northern catchments (NO1 and NO2) and the most southern catchment (CY). The EPI values
26 range between 1.11 and 1.2 for the 1 yr threshold, and 1.14 and 1.22 for the 5 yr threshold. For this
27 season, a similar variability is found for all catchments, except for CY, where the variability is
28 slightly larger than in the other catchments. For summer, the median is also greater than 1 for all the
29 catchments except for the two most southern catchments (CY and TR). These two catchments also
30 have a larger variability. In general, there are larger differences between and within the catchments
31 in summer than in winter.

1 In most catchments, and for both threshold (1 and 5 yrs), larger changes are expected for winter.
2 Only in the case of NO₂ the changes obtained for summer are larger than in winter. In the
3 catchment in LT, CZ1 and CZ2, larger changes are obtained for winter for the 1 yr level and for
4 summer for the 5 yr level. In both seasons and in most catchments, larger changes and variability
5 are obtained for the 5 yr level.

6 Comparing the changes obtained from the SDMs with the mean changes projected by the RCMs
7 (see Fig. 2), there is a general tendency that slightly smaller changes are estimated from the
8 uncorrected RCM projections. However, there are some significant differences. For example, for
9 NO₂ in winter and the 5 yr level, the uncorrected RCM projections point to a decrease of extreme
10 precipitation but the SDMs point to an increase. The opposite situation is obtained for CY for the
11 same season and 1 yr level. For this catchment (CY) in summer, there is also a rather large
12 difference between the changes estimated from the uncorrected RCM projections and the SDMs.
13 The largest difference between the uncorrected RCMs and downscaled results is obtained in CY.
14 The maximum difference is obtained in summer for the 5 yr level where the downscaled values lead
15 to a change 20% higher than the uncorrected RCMs. Excluding CY, the average difference of the
16 change between the downscaled and uncorrected series is small. For example, for the 1 yr level the
17 average difference is 0.013 for winter and 0.022 for summer. The smallest difference in both
18 seasons is obtained for the Danish catchment for which the difference is 0.003 in winter and 0.009
19 in summer. These overall results show that, in general, the SDMs do not modify the change
20 projected by the uncorrected RCMs significantly. Nonetheless, in some cases the use of some
21 downscaling methods might modify the magnitude of the change projected by the uncorrected
22 RCMs. The influence of the SDM used with respect to the difference between the change projected
23 by the uncorrected RCMs and the downscaled data is analysed in more detail in the next section.

FIGURE 2

27 Figure 2 does not differentiate between the variability due to the use of different SDMs and
28 different RCM-GCM simulations. The variance decomposition approach is used to assess each of
29 the sources of variance individually. Figure 3 shows the total variance decomposed in the variance
30 arising from the GCMs, RCMs, SDMs and the interaction terms for all catchments for the 1 and 5
31 yr levels and temporal aggregation of 1 day. For CY the results for the summer are not shown and

1 results for the winter do not include BCM and BCMV because EPI could not be calculated for a
2 large number of cases (due to the few rainy days in some of the RCM simulations).

3 As shown in Fig. 2, the variance for the 5 yr level is higher for all catchments and seasons than the
4 variance for the 1yr level. In summer, the variance tends to increase from north to south for the 5 yr
5 level, and to some extent also for the 1 yr level. This trend is not observed in winter. The larger
6 variance in the southern catchments for the 5 yr level may be partially caused by larger sampling
7 variance (smaller number of extreme events). Figure 3 shows that in most cases the variance due to
8 the RCM-GCM simulations is larger than the variance from the SDMs. However, the interaction
9 term is in both seasons and in most catchments similar or larger than the individual sources of
10 variance.

11 Figure 3 also shows the fractional percentage explained by $V(G)$, $V(R)$, and $V(S)$, such that the three
12 terms sum to 100%. The scaling of the percentages to obtain a total of 100% is needed because
13 some interaction terms are included in several factors. As already mentioned, the percentage
14 explained by the RCM-GCM simulations is in most cases larger than the percentage explained by
15 the SDMs. The only exception is TR for summer and PL for winter for the 1 yr level. However, in
16 all cases, the percentage explained by the SDMs is at least 30% of the total variance, which is
17 considerable. Similar results are obtained for winter and summer for the 1 and 5 yr levels. For both
18 seasons and return levels, there are no clear spatial patterns in the percentages. These results are in
19 agreement with the results obtained by Rätty et al. (2014). They carried out a similar variance
20 decomposition to study the variance arising from climate models and statistical downscaling
21 methods over northern and southern Europe. For northern Europe, they found that for the 70th and
22 higher precipitation percentiles, the climate models are the main source of variance and the variance
23 arising from the SDMs is at least 20% and the interaction term accounts for approximately 20%.
24 For southern Europe, the contribution of the SDMs is also at least 20%, but the variance arising
25 from the interaction term is higher (it ranges between 20 and 50% for all percentiles). In addition,
26 and also in agreement with the results shown here, Kidmose et al. (2013) found that for extreme
27 groundwater levels in a Danish catchment the variance arising from the ensemble of climate models
28 is higher than the variance arising from the SDMs, although only two downscaling methods were
29 considered. They also highlighted the importance of natural variability, which in their case was
30 higher than the variability related to climate models and downscaling methods. The results for
31 Norway (NO₂ and NO₁) are also in agreement with the results found by Lawrence and Haddeland

1 (2011). The influence of the SDMs is larger in the snow dominated catchment, NO1, than in the
2 rainfall dominated catchment, NO2.

3 In all cases the percentage of the variance explained by the RCMs is larger than the percentage
4 explained by the GCMs. For both return levels, in winter the average percentage explained by the
5 GCMs is approximately 20% while in summer it is approximately 15%. The smaller percentage for
6 the GCMs in the summer is due to the larger relative influence of both the RCMs and SDMs. This
7 is likely due to the fact that in Europe, extreme precipitation from convective storms occurs more
8 frequently during summer (e.g. Lenderink, 2010; Hofstra et al., 2009), and this has a larger
9 influence on the outputs from the RCMs and SDMs due to their higher spatial resolution. Several
10 studies have shown that the errors of the RCMs are larger in the representation of daily extreme
11 precipitation in summer over Europe (e.g. Frei et al., 2006; Fowler and Ekström, 2009).

12 The results of the variance decomposition obtained for aggregation levels larger than 1 day (not
13 shown) point towards a smaller total variance. For these temporal aggregations, the main source of
14 variation is also the RCM-GCMs, although the percentage explained by SDMs is slightly larger
15 than for the 1 day aggregation. The decrease in total variance and in the percentage explained by
16 RCM-GCMs mainly reflects that the model outputs being more similar for larger temporal
17 aggregations. The results from the variance decomposition highlight the need for considering both a
18 range of SDMs and an ensemble of RCMs driven by different GCMs for assessing the uncertainty
19 in the projection of changes in extreme precipitation.

20 FIGURE 3

21 4.1.2. Extreme precipitation index for three selected catchments

22 The previous section summarizes the main results regarding the expected changes in extreme
23 precipitation when considering all the RCMs and SDMs. This section focuses on the differences
24 between the statistical downscaling methods. For this purpose, three catchments have been selected:
25 NO2, DE, and TR (distributed north to south and with different precipitation patterns). Figure 4
26 shows the median, 25th, and 75th quantile of EPI for each SDM for the three catchments for the 1 yr
27 level and a temporal aggregation of 1 day.

28 In NO2, for both seasons, the SDMs based on BC show a lower EPI than the methods based on
29 CFs. In winter, all the CF methods point towards an increase in extreme precipitation, although
30 some of the BC methods show a decrease for some RCMs. In summer, all methods point to an
31 increase except XDS, which produces a small EPI and a large variability. There are several factors

1 which may contribute to these differences. As this region is projected to generally have an increase
2 in winter precipitation, use of change factor methods that do not correct for changes in the number
3 of wet days will automatically produce higher values for extreme precipitation in winter. If this
4 precipitation increase is, however, also associated with a change in storm patterns, such that the
5 increase simply reflects an increase in wet days rather than wet day extremes, then this difference
6 would be reflected in the results for the BC methods.

7 In DE, all the SDMs lead to similar median values except the BCMV in winter and CFM in
8 summer. The differences between BCMV and the other two BC methods are due to some RCMs
9 leading to very large changes when they are downscaled with BCMV, e.g. for RCA-ECHAM5, the
10 values of EPI are 1.18 for BCM, 1.16 for BCQM and 1.63 for BCMV. This large value of EPI is
11 caused by unexpectedly large precipitation intensities obtained from the non-linear transformation
12 in BCMV, which is one of the disadvantages of this method (see Table 3). For the BCMV method
13 two events of 55 and 60 mm/day are obtained while the largest events for the two other BC methods
14 are below 40 mm/day (for the control period all the events are lower than 30 mm/day).

15 CFM leads to the lowest value of EPI obtained in summer. This is also the case for all the other
16 catchments considered in this study except NO2 and Yermasoyia in Cyprus (results not shown). It
17 indicates that mean precipitation is likely to increase less than the more extreme precipitation
18 intensities. In addition, it illustrates that the CFM method is not suitable for regions where the
19 expected changes in extreme precipitation are different than the changes in mean precipitation.

20 In TR, the results of the SDMs vary more than in DE and NO2. For this catchment, CFM leads to
21 the lowest EPI in both seasons, which indicates a lower increase in mean precipitation than in
22 extreme precipitation, as in DE. In summer, all SDMs point to a decrease of extreme precipitation
23 except BCM and BCMV, which do not show a change in extreme precipitation. These two methods
24 show the largest variability for both winter and summer. The high variability for these two methods
25 is due to the same issue identified in CY, i.e. only a few rainy days in the RCM simulations, the
26 annual percentage of rainy days ranges between 12% and 28% rainy days.

27 For all catchments and both seasons, very similar results are obtained for CFQM and CFQP. This is
28 expected since the main difference between the two methods is the treatment of wet day frequency.
29 This is expected to have a minor impact, except for TR in the summer, where there are only very
30 few rainy days during the summer period. This implies that in some cases all the rainy days are
31 included in the selection of extreme events. Hence, the change in the number of wet days may have

1 an effect on the changes in extreme precipitation. Similar results to those illustrated in Fig. 4 were
2 also obtained for the 5 yr level (results not shown).

3 The results for the three catchments show that there is not a clear tendency in the differences
4 between CF and BC methods. In addition, there is no evidence that methods that are based on the
5 same statistics for the correction (e.g. BCM and CFM or BCMV and CFMV) will lead to similar
6 results. Hence, it is not possible to generalize the results with respect to the use of SDM. This result
7 contrasts with the findings in Hanel et al. (2013) for low flows in the Czech Republic. They found
8 that, in general, the SDMs which account for changes in variance (such as BCMV and CFMV) led
9 to larger changes in runoff. In addition, they also found larger changes in runoff for BC than for CF
10 methods.

11 The EPI estimated using the uncorrected RCMs can be used as a reference to assess whether the
12 downscaled data preserves the changes projected by the RCMs and the differences depending on the
13 SDM. In the case of NO₂, the EPI estimated using the uncorrected RCMs lie in between the values
14 from the BC and CF methods. The downscaling method that shows the closest agreement with the
15 changes projected by the RCMs is BCQM. Overall for the three catchments and both seasons this
16 method is the one that shows values of EPI closest to the ones estimated from the uncorrected
17 RCMs. This points towards the suitability of this method to downscale extreme precipitation as it
18 corrects the properties of interest for representing extreme precipitation. On the other hand, EPI
19 obtained from CFM tend to produce the largest deviations from the EPI of the uncorrected RCMs
20 (except in the case of TR in summer), which again shows that this method is not suitable for
21 projecting changes in extreme precipitation. In addition, problems of producing unrealistic extreme
22 precipitation values with some of the methods, such as BCM and BCMV in TR in summer, XDS in
23 TR in winter and NO₂ in summer are clearly seen when comparing their EPI values with those
24 obtained from the uncorrected RCMs. The above examples illustrate that some SDMs are better
25 suited for downscaling extreme precipitation and some SDMs are less robust with respect to
26 downscaling various precipitation patterns.

27 FIGURE 4

28 Figure 5 analyses the eight SDMs for the three catchments for two temporal aggregations: 1 and 30
29 days. In general, the variability in EPI in the RCM ensemble decreases with increasing temporal
30 aggregation, except for a few cases, e.g. XDS in NO₂ and BCM for DE in summer. There is no
31 general indication that EPI either increases or decreases with increasing temporal aggregation.

1 In NO₂, EPI is larger for a temporal aggregation of 30 days for BCM, BCMV and BCQM, and it is
2 lower for the CF methods and XDS for summer. In winter, EPI for BCM, BCMV and BCQM is
3 also slightly larger for a temporal aggregation of 30 days (in the case of BCM and BCMV, this
4 means a smaller reduction of extreme precipitation). In DE, most methods show a lower EPI for 30
5 days except CFM in summer and CFM, CFMV and XDS in winter. Similarly, in TR all the methods
6 show lower EPI for 30 days except for CFM, XDS and CFQM in summer. For all catchments, the
7 results of the SDMs at 30 days temporal aggregation are more similar than for 1 day aggregation.

8 In most cases, EPI at 1 and 30 days are not considerably different and show the same signal (except
9 in the case of TR for BCM and BCMV for both seasons and BCQM in winter). As for the 1 day
10 aggregation, the results with temporal aggregation of 30 days do not allow general conclusions with
11 respect to the use of SDM.

12 **FIGURE 5**

13 **4.2. Comparison of observations and bias corrected RCMs for the control period**

14 The previous section focuses on the analysis of the expected changes in extreme precipitation. This
15 section uses EPI to compare the results from the BC methods for the control period and the
16 observations. This allows us to evaluate how well the different BC methods correct extreme
17 precipitation from the RCMs. As in the previous section, a summary of the results found for all the
18 catchments is first presented, followed by a more detailed analysis of the results found for each BC
19 method for three of the catchments. **It must be noted that this comparison of the results for the
20 control period does not provide a validation of the downscaling methods. The data used to
21 downscale the RCMs for the control period is the same as the data used for the calibration of these
22 methods. Nonetheless, it should be noted that the validation of downscaling methods is crucial and
23 relevant for assessing how well we can estimate changes in extreme precipitation. However, the
24 validation of SDMs is challenging as it requires either observational data that have different
25 properties to enable assessing whether the downscaling methods can be used to project climate
26 changes (e.g. Refsgaard et al. 2014; Teutschbein and Seibert, 2012) or, alternatively, the use of
27 pseudo-realities (e.g. Räisänen and Rätty 2013; Vrac et al. 2007; Maraun et al., 2015). If the
28 observational data do not show pronounced changes in extremes, then the results of the validation
29 analyses are questionable with respect to the suitability of the methods for use in climate change
30 analyses. There is, thus, a clear need for further research on validation methods for SDMs and will
31 not be addressed in this paper.**

1 For BE, CY, CZ2, DK, and PL, the control period considered for the RCMs does not fully overlap
2 with the observation period. In the case of DK, for example, there is only an overlap of 2 yrs. The
3 use of different periods assumes that the statistics are stationary between the periods. However,
4 some of the disagreements between the observations and bias corrected results may well be due to
5 non-stationary statistics between the two periods.

6 **4.2.1. Extreme precipitation index for all catchments**

7 Figure 6 shows EPI estimated using the observations and the bias corrected RCM. In this figure
8 (and the rest of the figures in this section), a value of 1 indicates that there is no difference between
9 the extreme value statistics from the observations and the bias corrected RCM. A value greater
10 (less) than one indicates that the bias corrected RCM overestimates (underestimates) extreme
11 precipitation. It must be noted that for the catchments LT and TR there is a perfect overlap between
12 the time period of the observations and RCMs, while for the other catchments the observation
13 period includes the RCM period or there is only a partly overlap between the time period of the
14 observations and RCMs (see Table 1 for details).

15

FIGURE 6

16 For extreme winter precipitation there is no clear tendency across catchments for under- or
17 overestimation with the bias corrected data. The catchments that have the largest underestimation
18 are for the most northern and southern catchments (NO2, NO1, DK and CY), whereas LT, BE and
19 PL have the largest overestimation. For extreme summer precipitation, there is a pronounced
20 underestimation for a number of catchments. The three most northern catchments (NO2, NO1, and
21 DK) show the lowest mean bias based on the median values for all downscaled projections. The
22 most southern catchment (CY) has the largest underestimation of extreme summer precipitation.
23 Both the median and variance of EPI depend on the catchment and the season. For example, the
24 bias corrected data for LT, BE and PL tend to overestimate extreme precipitation in winter, but
25 underestimate this in summer. CZ1 in winter and NO2 in summer are the catchments that lead to the
26 median closest to 1. The largest variability is found for PL in winter and TR and CY in summer.

27 The comparison of the error in the RCMs before and after bias correction shows that, in general, the
28 error after bias correction is smaller than before bias correction. This shows that the BC methods
29 improve the representation of extremes. However, in a few cases the error of the RCMs before bias
30 correction is smaller than after bias correction. This is because some of the RCMs result in large
31 errors after bias correction. For example for BE in winter with the HadRM3Q3-HadCM3Q3 model,

1 values of 1.18 for BCM, 1.37 for BCMV, 1.24 for BCQM, and 1.23 for XDS are obtained, while a
2 value of 0.98 is obtained from the uncorrected data. In fact, the average over all the RCMs shows
3 that none of the downscaling methods improves the results of the uncorrected RCMs for this
4 catchment. A similar result is obtained for the DE catchment. In the summer period, the results after
5 bias correction for all the downscaling methods in the LT catchment show larger differences
6 compared to the observations than the uncorrected RCMs. In both seasons, these results (error of
7 the RCMs before bias correction is smaller than after bias correction) are obtained for catchments
8 where the RCMs have the lowest error in representing observed extreme precipitation (i.e. EPI
9 closer to 1). This indicates that if the agreement between the observations and RCMs is high, the
10 downscaling methods considered in this study are not able to improve it. The next section describes
11 in more detail the difference between EPI of the uncorrected RCMs and the downscaled series for
12 each bias correction method.

13 **4.2.2. Extreme precipitation index for each bias correction method for three** 14 **selected catchments**

15 Figure 7 shows the results of the three BC methods and XDS for NO₂, DE, and TR for the 1 yr
16 level and 1 day temporal aggregation. The performance of each method varies depending on the
17 season and catchment. For example, BCM overestimates extremes in NO₂ in winter and TR in
18 summer and underestimates them in NO₂ in summer and TR in winter. In DE, BCM performs
19 equally well as BCMV. This illustrates that simple BC methods can, in some cases perform
20 similarly or better than more advanced methods. In the catchments considered in this study, there is
21 no clear relationship between the performance of the BC methods and the precipitation regime for
22 the catchments.

23 In winter, the errors obtained for DE are smaller than in the other two catchments. EPI ranges from
24 an underestimation of 4% (EPI equal to 0.96) for BCM and BCMV, to an overestimation of
25 approximately 6% for BCQM and XDS. For this catchment and both seasons, BCM and BCMV
26 lead to better results than BCQM and XDS. In summer, the errors in NO₂ are smaller than in the
27 other two catchments. For this catchment and this season, XDS is the method that leads to the
28 smallest error and variability.

29 The largest errors and variability in the results are found for the TR catchment in both seasons. For
30 this catchment and in the winter period, the median of all methods underestimate extremes except
31 XDS, while in summer BCM and BCMV overestimate extremes and the other two methods

1 underestimate. A very large variability is obtained for BCM and BCMV in summer (the 25th and
2 75th percentiles range from 0.4 to 1.5).

3 Comparison of the results of the SDMs with EPI obtained from the uncorrected RCMs shows that
4 in the case of NO₂ all the SDMs clearly agree better with the observations. But for the other two
5 catchments, the results depend on the downscaling method. In DE, BCM and BCMV lead to better
6 results than the other two methods for both seasons. In the TR catchment, BCQM leads to the best
7 result in winter but not in summer, where BCMV produces the best result. Even though the results
8 depend on the catchment analysed, the BCM is the method leads to the least improvements in most
9 cases compared to the results of the uncorrected RCM. This is in agreement with the main
10 conclusion from the validation study carried out by Teutschbein and Seibert (2012). They
11 concluded that the linear bias correction (equivalent to the BCM method used here) together with
12 the delta-change method (equivalent to the CFM used here) are less reliable than other more
13 complex methods. Similarly, the cross-validation study carried out by Rätty et al. (2014) showed
14 that the linear bias correction method tends to perform more poorly than the other more complex
15 bias correction methods, especially for high percentiles (between 75th and 97th percentile) in
16 southern Europe and between the 50th and 70th percentile in northern Europe. Nonetheless, it should
17 be noted that even if in some cases it is possible to identify a method that performs better than
18 others it might not be possible to reject the hypothesis that all SDMs perform equally well
19 (Wetterhall et al., 2012). This points towards the advantage of using an ensemble of SDMs to
20 represent the uncertainty related to the statistical downscaling.

21 FIGURE 7

22
23 The results from Figure 7 indicate that the bias correction methods do not in all cases improve the
24 time series from the RCMs. This must be tested for each application. Figure 8 shows the error of
25 each BC method for two temporal aggregations, 1 and 30 days, for the 1 yr level. In general, the
26 performance of the BC methods for the winter period improves for large temporal aggregation
27 (except for XDS in TR). However, in summer this is not the case. For this season, the difference
28 between the results for 1 and 30 day aggregations depends on the catchment and the method. In
29 NO₂, the results for 1 day are better than for 30 days for BCQM and XDS, although the reverse is
30 true for TR. In DE, the results for 1 day are better than for 30 days for all the methods except XDS.

1 As shown in Fig. 7, TR has the largest variability for 30 days followed by NO2 for both seasons.
2 The results for DE appear to be the least dependent on the temporal aggregation. This may be the
3 result of spatially averaging the observations from 43 stations to derive the catchment precipitation.
4 For such a large basin (6171 km², see Table 1), this may simultaneously lead to temporally-
5 averaged precipitation values from the gauged nested sub-catchments. In all cases, the variability
6 for 30 days is smaller than for 1 day, indicating that the RCMs lead to more similar results for large
7 temporal aggregations.

8 FIGURE 8

9 **5. Summary and conclusions**

10 This study analyses the expected changes in extreme precipitation in eleven European catchments.
11 It focuses on the variability in the changes arising from the use of different statistical downscaling
12 methods as well as different RCM-GCM simulations. Fifteen RCMs driven by six GCMs are
13 downscaled using eight statistical downscaling methods. The statistical downscaling methods rely
14 on different assumptions and different RCM outputs. The outputs from all the statistical
15 downscaling methods are analysed using an extreme precipitation index.

16 Extreme precipitation is expected to increase in most catchments in both winter and summer. A
17 decrease in extreme precipitation is only expected for both winter and summer in CY and for
18 summer in TR. In most catchments, larger changes are expected in winter than in summer.
19 Additionally, in all cases, larger increases and larger variability in the results are obtained for **the**
20 higher return level, **5 years**.

21 In most catchments and for both winter and summer, the RCM-GCM projections are the main
22 source of variability in the results when compared to the differences between SDMs, although
23 variability due to the SDMs explains at least 30% of the total variance in all cases. Additionally, in
24 all cases, the RCMs represent a larger percentage of the total variability than the GCMs, especially
25 in summer. For this season, the total variance tends to be higher for the most southern catchments.

26 In general, the eight statistical downscaling methods agree on the direction of the change but not the
27 magnitude of the change. It is not possible to draw general conclusions regarding differences
28 between the downscaling methods, as the differences depend on the physical geographical
29 characteristics of the catchment and the season analysed. For example, for NO2 the bias correction
30 methods lead to lower changes than the change factor methods, but this is not the case for the other
31 catchments. A common result for all catchments except NO2 and CY is that the CFM method leads

1 to the smallest increase of extreme precipitation in summer. This indicates that this method is not
2 suitable for regions where the expected changes in extreme precipitation differ from the changes in
3 mean precipitation. The changes obtained for different temporal aggregations also depend on the
4 physical geographical characteristics of the catchment and season analysed, i.e. there is no general
5 tendency for an increase or decrease in the index with increasing temporal aggregation.

6 Overall, the bias correction methods improve the representation of extreme precipitation, as
7 compared with the uncorrected RCM outputs. However, the bias corrected time series tend to
8 underestimate extreme precipitation. The magnitude of the errors depends on the catchment and
9 season analysed. For example, the results of the bias correction of mean are worse than the other
10 methods for the NO2 but not for the other catchments. There is no clear relationship between the
11 performance of the bias correction methods and the precipitation regime of the catchment. There is
12 also no clear indication of an increase or decrease in the error with increasing temporal aggregation.

13 The results from the statistical downscaling methods have been compared with the extreme
14 precipitation obtained from the uncorrected RCMs. Although the results depend on the catchment
15 and season as in the other comparisons discussed before, some overall conclusions can be extracted
16 from this comparison. Regarding the comparison of the change in extreme precipitation projected
17 by the uncorrected RCMs and the downscaled series, the SDM that showed the smallest differences
18 relative to the RCM projections is the BCQM method, while the method that led to the largest
19 differences is the CFM method. These differences between the methods are more pronounced for
20 the summer period. From the comparison of the SDMs and the uncorrected RCMs in representing
21 the current period it was found that in general the BCM method fails in more cases than the other
22 SDMs in improving the representation of extreme precipitation from the uncorrected RCMs.

23 From the results of all these comparisons, it is possible to draw some general recommendations
24 when selecting SDMs from the ones considered here for downscaling extreme precipitation.
25 Downscaling methods that do not explicitly correct or take into account changes in extreme
26 precipitation may lead to different climate change signals than the ones projected by the RCMs and
27 should not be used. In this study, this occurs mainly with CFM. In addition, some methods fail to
28 correct the errors in the RCMs in representing extreme precipitation. In this study, this occurred in
29 more cases when using BCM than with the other methods. Finally, in catchments with long dry
30 periods the BCM, BCMV, CFM, CFMV, and CFQM methods produce unrealistic results and
31 should not be used (or should be configured differently than done in this study with respect to

1 describing the seasonal patterns). BCMV may also lead to unrealistic results in other catchments as
2 seen in the case of DE. The ability of the downscaling methods to improve the representation of
3 extreme precipitation from the RCMs and to preserve the climate change signal should be assessed
4 for each case study in order to select the most suitable SDMs.

5 This study illustrates that there is a large variability in the changes estimated from different
6 statistical downscaling methods and RCMs. It also shows that the differences between the methods
7 and the performance of the bias correction methods depend on the catchment studied. Hence, for a
8 specific case study, the selection of a suitable statistical downscaling method may depend on the
9 physical geographical characteristics of the catchment. However, we recommend the use of a set of
10 statistical downscaling methods as well as an ensemble of climate model projections. The selection
11 of statistical downscaling methods should include: methods that are able to project changes in
12 extreme precipitation if they are expected to be different from other precipitation properties;
13 methods based on different underlying assumptions, for example BC and CF methods; and methods
14 that use different outputs from the RCMs as, for example, XDS, CF or BC methods including mean
15 and variance of precipitation, and methods including a range of quantiles.

16 **Acknowledgements**

17 This work was carried out as part of the COST Action ES0901 “European Procedures for Flood Frequency
18 Estimation (FloodFreq)” (<http://www.cost-floodfreq.eu/>). Maria A. Sunyer and Henrik Madsen were
19 supported by the Danish Council for Strategic Research as part of the project RiskChange. Deborah
20 Lawrence acknowledges support from NVE for the internal research project ‘Climate change and future
21 floods’. Patrick Willems was supported by research projects for the Fund for Scientific Research (F.W.O.) -
22 Flanders, Flanders Hydraulics Research and the Flemish Environment Agency. Klaus Vormoor
23 acknowledges funding from the Helmholtz graduate research school GeoSim. Marta Martínková was
24 financially supported by Internal Grant Agency of the Faculty of Environmental Sciences, CULS Prague
25 (00000869/2013). Gerd Bürger was funded by the Austrian Climate and Energy Fund as part of the Austrian
26 Climate Research Programme. Ismail Yucel was supported by TÜBİTAK project number 110Y036. The data
27 from the RCMs used in this work was funded by the EU FP6 Integrated Project ENSEMBLES contract
28 number 05539 (<http://ensembles-eu.metoffice.com>), whose support is gratefully acknowledged.

29 **References**

30 Boé, J., Terray, L., Habets, F. and Martin, E.: Statistical and dynamical downscaling of the Seine basin
31 climate for hydro-meteorological studies, *Int. J. Climatol.*, 27(12), 1643–1655, doi:10.1002/joc.1602, 2007.

1 Bürger, G. and Chen, Y.: Regression-based downscaling of spatial variability for hydrologic applications, *J.*
2 *Hydrol.*, 311(1-4), 299–317, doi:10.1016/j.jhydrol.2005.01.025, 2005.

3 Bürger, G., Reusser, D. and Kneis, D.: Early flood warnings from empirical (expanded) downscaling of the
4 full ECMWF Ensemble Prediction System, *Water Resour. Res.*, 45(10), n/a–n/a,
5 doi:10.1029/2009WR007779, 2009.

6 Bürger, G., Murdock, T.Q., Werner, A.T., Sobie, S.R. and Cannon, A.J.: Downscaling Extremes—An
7 Intercomparison of Multiple Statistical Methods for Present Climate, *J. Clim.*, 25, 4366–4388, doi:
8 10.1175/JCLI-D-11-00408.1, 2012.

9 Bürger, G., Sobie, S. R., Cannon, A. J., Werner, A. T. and Murdock, T. Q.: Downscaling Extremes: An
10 Intercomparison of Multiple Methods for Future Climate, *J. Clim.*, 26(10), 3429–3449, doi:10.1175/JCLI-D-
11 12-00249.1, 2013.

12 Christensen, J. H. and Christensen, O. B.: Climate modelling: Severe summertime flooding in Europe.,
13 *Nature*, 421(6925), 805–6, doi:10.1038/421805a, 2003.

14 Danish Meteorological Institute (DMI): Climate Grid Denmark. Dataset for use in research and education.
15 Daily and monthly values 1989-2010 10x10 km observed precipitation 20x20 km temperature, potential
16 evaporation (Makkink), wind speed, global radiation, Technical Report 12-10, report available at
17 <http://beta.dmi.dk/fileadmin/Rapporter/TR/tr12-10.pdf> (last access: 3 June 2014), 2012.

18 Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., Jacob, D., Kjellström, E.,
19 Castro, M. and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing
20 uncertainties in model projections, *Clim. Change*, 81(S1), 53–70, doi:10.1007/s10584-006-9228-x, 2007.

21 Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C. M., Jacob, D., Lenderink, G. and Christensen, O. B.:
22 The spread amongst ENSEMBLES regional scenarios: regional climate models, driving general circulation
23 models and interannual variability, *Clim. Dyn.*, 38, 951–964, doi:10.1007/s00382-011-1053-x, 2012.

24 Dobler, C., Bürger, G. and Stötter, J.: Assessment of climate change impacts on flood hazard potential in the
25 Alpine Lech watershed, *J. Hydrol.*, 460-461, 29–39, doi:10.1016/j.jhydrol.2012.06.027, 2012.

26 Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections
27 for use by impact models: Evaluation on the present climate, *J. Geophys. Res.*, 116(D16), D16106,
28 doi:10.1029/2011JD015934, 2011.

29 Fowler, H. J., Blenkinsop, S. and Tebaldi, C.: Linking climate change modelling to impacts studies : recent
30 advances in downscaling techniques for hydrological modelling, *Int. J. Climatol.*, 27, 1547–1578,
31 doi:10.1002/joc1556, 2007.

1 Fowler, H. J. and Ekström, M.: Multi-model ensemble estimates of climate change impacts on UK seasonal
2 precipitation extremes, *Int. J. Climatol.*, 29(3), 385–416, doi:10.1002/joc.1827, 2009.

3 Frei, C., Schöll, R., Fukutome, S., Schmidli, J. and Vidale, P. L.: Future change of precipitation extremes in
4 Europe: Intercomparison of scenarios from regional climate models, *J. Geophys. Res.*, 111, D06105,
5 doi:10.1029/2005JD005965, 2006.

6 Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen-Skaugen, T.: Technical Note: Downscaling
7 RCM precipitation to the station scale using statistical transformations – a comparison of methods, *Hydrol.*
8 *Earth Syst. Sci.*, 16(9), 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

9 Gudmundsson, L.: qmap: Statistical transformations for post-processing climate model output. R package
10 version 1.0-2, technical note available at: <http://cran.r-project.org/web/packages/qmap/qmap.pdf> (last access:
11 3 June 2014), 2014.

12 Hanel, M., Mrkvičková, M., Máca, P., Vizina, A. and Pech, P.: Evaluation of Simple Statistical Downscaling
13 Methods for Monthly Regional Climate Model Simulations with Respect to the Estimated Changes in Runoff
14 in the Czech Republic, *Water Resour. Manag.*, 27, 5261–5279, doi:10.1007/s11269-013-0466-1, 2013.

15 Ho, C. K., Stephenson, D. B., Collins, M., Ferro, C. a. T. and Brown, S. J.: Calibration Strategies: A Source
16 of Additional Uncertainty in Climate Change Projections, *Bull. Am. Meteorol. Soc.*, 93(1), 21–26,
17 doi:10.1175/2011BAMS3110.1, 2012.

18 Hofstra, N., Haylock, M., New, M. and Jones, P. D.: Testing E-OBS European high-resolution gridded data
19 set of daily precipitation and surface temperature, *J. Geophys. Res.*, 114, D21101,
20 doi:10.1029/2009JD011799, 2009.

21 Hundecha, Y., Sunyer, M.A., Lawrence, D., Willems, P., Bürger, G., Kriaučiūnienė, J., Loukas, A.,
22 Martinkova, M., Osuch, M., Vormoor, K., Yücel, I., Madsen, H.: Effect of downscaling climate data on
23 indices of extreme river flow: A comparative study across Europe, in preparation.

24 IPCC: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A
25 Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, edited by C.
26 B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K.
27 Plattner, S. K. Allen, M. Tignor, and P. M. Midgley, Cambridge University Press, Cambridge, UK, and New
28 York, NY, USA., 2012.

29 Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A.,
30 Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A.,
31 Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E.,
32 Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D.,
33 Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R.,

- 1 Weber, B. and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European
2 impact research, *Reg. Environ. Chang.*, 14(2), 563–578, doi:10.1007/s10113-013-0499-2, 2013.
- 3 Kidmose, J., Refsgaard, J. C., Trolborg, L., Seaby, L. P. and Escrivà M. M.: Climate change impact on
4 groundwater levels: ensemble modelling of extreme values, *Hydrol. Earth Syst. Sci.*, 17, 1619–1634,
5 doi:10.5194/hess-17-1619-2013, 2013.
- 6 Kendon, E. J., Rowell, D. P., Jones, R. G. and Buonomo, E.: Robustness of Future Changes in Local
7 Precipitation Extremes, *J. Clim.*, 21(17), 4280–4297, doi:10.1175/2008JCLI2082.1, 2008.
- 8 Knutti, R.: The end of model democracy?, *Clim. Change*, 102(3-4), 395–404, doi:10.1007/s10584-010-9800-
9 2, 2010.
- 10 Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. and Meehl, G. a.: Challenges in Combining Projections from
11 Multiple Climate Models, *J. Clim.*, 23, 2739–2758, doi:10.1175/2009JCLI3361.1, 2010.
- 12 Lawrence, D. and Haddeland, I.: Uncertainty in hydrological modelling of climate change impacts in four
13 Norwegian catchments, *Hydrol. Res.*, 42(6), 457, doi:10.2166/nh.2011.010, 2011.
- 14 Leander, R. and Buishand, T. A.: Resampling of regional climate model output for the simulation of extreme
15 river flows, *J. Hydrol.*, 332(3-4), 487–496, doi:10.1016/j.jhydrol.2006.08.006, 2007.
- 16 Leander, R., Buishand, T. A., van den Hurk, B. J. J. M. and de Wit, M. J. M.: Estimated changes in flood
17 quantiles of the river Meuse from resampling of regional climate model output, *J. Hydrol.*, 351(3-4), 331–
18 343, doi:10.1016/j.jhydrol.2007.12.020, 2008.
- 19 Lenderink, G.: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate
20 model simulations, *Clim. Res.*, 44, 151–166, doi:10.3354/cr00946, 2010.
- 21 Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust,
22 H. W., Sauter, T., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M. and
23 Thiele-Eich, I.: Precipitation downscaling under climate change. Recent developments to bridge the gap
24 between dynamical models and the end user, *Rev. Geophys.*, 48(3), RG3003, 2010.
- 25 Maraun, D., Widmann, M., Benestad, R., Kotlarski, S., Huth, R., Hertig, E., Wibig, J. and Gutierrez, J.:
26 VALUE - Validating and Integrating Downscaling Methods for Climate Change Research, EGU Gen.
27 Assem. Conf. Abstr., 15, 12041, 2013.
- 28 Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R.
29 and Wilcke, R.A.I.: VALUE: A framework to validate downscaling approaches for climate change studies.
30 *Earth's Future*. doi: 10.1002/2014EF000259, 2015.
- 31 Ntegeka, V., Baguis, P., Roulin, E. and Willems, P.: Developing tailored climate change scenarios for
32 hydrological impact assessments, *J. Hydrol.*, 508, 307–321, doi:10.1016/j.jhydrol.2013.11.001, 2014.

- 1 Olsson, J., Berggren, K., Olofsson, M. and Viklander, M.: Applying climate model precipitation scenarios
2 for urban hydrological assessment: A case study in Kalmar City, Sweden, *Atmos. Res.*, 92, 364–375,
3 doi:10.1016/j.atmosres.2009.01.015, 2009.
- 4 Piani, C., Haerter, J. O. and Coppola, E.: Statistical bias correction for daily precipitation in regional climate
5 models over Europe, *Theor. Appl. Climatol.*, 99, 187–192, doi:10.1007/s00704-009-0134-9, 2010.
- 6 Prudhomme, C., Reynard, N. and Crooks, S.: Downscaling of global climate models for flood frequency
7 analysis: where are we now?, *Hydrol. Process.*, 16(6), 1137–1150, doi:10.1002/hyp.1054, 2002.
- 8 Räisänen, J. and Rätty, O.: Projections of daily mean temperature variability in the future: cross-validation
9 tests with ENSEMBLES regional climate simulations, *Clim. Dyn.*, 41, 1553–1568, doi:10.1007/s00382-012-
10 1515-9, 2013.
- 11 Rana, A., Foster, K., Bosshard, T., Olsson, J. and Bengtsson, L.: Impact of climate change on rainfall over
12 Mumbai using Distribution-based Scaling of Global Climate Model projections, *J. Hydrol.: Reg. Stud.*, 1,
13 107–128, doi: 10.1016/j.ejrh.2014.06.005, 2014.
- 14 Rätty, O., Räisänen, J. and Ylhäisi, J.S.: Evaluation of delta change and bias correction methods for future
15 daily precipitation: intermodel cross-validation using ENSEMBLES simulations, *Clim. Dyn.*, 42, 2287–
16 2303, doi: 10.1007/s00382-014-2130-8, 2014.
- 17 Rummukainen, M.: Methods of statistical downscaling of GCM simulations, *Reports Meteorology and*
18 *Climatology* 80, Tech. rep., Swedish Meteorological and Hydrological Institute, SE-601 76 Norrköping,
19 Sweden, 1997.
- 20 Šercl, P.: Assessment of methods for area precipitation estimates, *Meteorological Bulletin*, 61(2), 2008.
- 21 Sunyer, M. A., Madsen, H. and Ang, P. H.: A comparison of different regional climate models and statistical
22 downscaling methods for extreme rainfall estimation under climate change, *Atmos. Res.*, 103, 119–128,
23 doi:10.1016/j.atmosres.2011.06.011, 2012.
- 24 Sunyer, M.A., Madsen, H., Rosbjerg, D., Arnbjerg-Nielsen, K.: A Bayesian approach for uncertainty
25 quantification of extreme precipitation projections including climate model interdependency and non-
26 stationary bias, *J. Climate*, 27, 7113–7132, doi:10.1175/JCLI-D-13-00589.1, 2014.
- 27 Taye, M. T., Ntegeka, V., Ogiramoi, N. P. and Willems, P.: Assessment of climate change impact on
28 hydrological extremes in two source regions of the Nile River Basin, *Hydrol. Earth Syst. Sci.*, 15, 209–222,
29 doi:10.5194/hess-15-209-2011, 2011.
- 30 Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections., *Philos.*
31 *Trans. A. Math. Phys. Eng. Sci.*, 365(1857), 2053–2075, doi:10.1098/rsta.2007.2076, 2007.

1 Teutschbein, C. and Seibert, J.: Is bias correction of Regional Climate Model (RCM) simulations possible for
2 non-stationary conditions?, *Hydrol. Earth Syst. Sci. Discuss.*, 9(11), 12765–12795, doi:10.5194/hessd-9-
3 12765-2012, 2012.

4 Tveito, O. E., Bjørndal, I., Skjelvåg, A. O. and Aune, B.: A GIS-based agro-ecological decision system based
5 on gridded climatology, *Meteorol. Appl.*, 12(1), 57–68, doi:10.1017/S1350482705001490, 2005.

6 Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson J. K.,
7 Haseler J., Hernandez A., Kelly G. A., Li X., Onogi K., Saarinen S., Sokka N., Allan R. P., Andersson E.,
8 Arpe K., Balmaseda M. A., Beljaars A. C. M., Van De Berg L., Bidlot J., Bormann N., Caires S., Chevallier
9 F., Dethof A., Dragosavac M., Fisher M., Fuentes M., Hagemann S., Hólm E., Hoskins B. J., Isaksen L.,
10 Janssen P. A. E. M., Jenne R., McNally A. P., Mahfouf J.-F., Morcrette J.-J., Rayner N. A., Saunders R. W.,
11 Simon P., Sterl A., Trenberth K. E., Untch A., Vasiljevic D., Viterbo P. and Woollen, J.: The ERA-40 re-
12 analysis, *Q. J. Roy. Meteor. Soc.*, 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

13 van der Linden P., and Mitchell J.F.B. (eds.): ENSEMBLES: Climate Change and its Impacts: Summary of
14 research and results from the ENSEMBLES project. Met Office Hadley Centre, FitzRoy Road, Exeter EX1
15 3PB, UK. 160pp, 2009. Technical report available at: [http://ensembles-](http://ensembles-eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf)
16 [eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf](http://ensembles-eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf) (last access: 3 June 2014)

17 Vansteenkiste, Th., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F. and Willems, P.:
18 Intercomparison of hydrological model structures and calibration approaches in climate scenario impact
19 projections. *J. Hydrol.*, 519, 743–755, doi: 10.1016/j.jhydrol.2014.07.062, 2014.

20 Vrac, M., Stein, M. L., Hayhoe, K. and Liang, X.-Z.: A general method for validating statistical downscaling
21 methods under future climate change, *Geophys. Res. Lett.*, 34(18), L18701, doi:10.1029/2007GL030295,
22 2007.

23 Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P. and Mearns, L. O.: Guidelines for Use of
24 Climate Scenarios Developed from Statistical Downscaling Methods., 2004.

25 Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow
26 scenarios for the River Thames, UK, *Water Resour. Res.*, 42(2), n/a–n/a, doi:10.1029/2005WR004065, 2006.

27 Willems, P. and Vrac, M.: Statistical precipitation downscaling for small-scale hydrological impact
28 investigations of climate change, *J. Hydrol.*, 402, 193–205, doi:10.1016/j.jhydrol.2011.02.030, 2011.

29 Willems, P.: Revision of urban drainage design rules after assessment of climate change impacts on
30 precipitation extremes at Uccle, Belgium, *J. Hydrol.*, 496, 166–177, doi: 10.1016/j.jhydrol.2013.05.037,
31 2013.

32

1 **Appendix A - Verification of matrix reconstruction approach**

2 **A.1 Comparison of results using 2 and 3 sources of variance**

3 This verification approach assesses the influence of the matrix reconstruction procedure on the
4 percentage of the total variance explained by climate models (influence of GCM-RCM simulations)
5 and SDMs. For this purpose, the variance decomposition approach has been applied considering
6 two sources of uncertainty: SDMs and climate models (the 15 RCM-GCM simulations). In the case
7 of two sources of variance, there is no need to reconstruct the matrix.

8 Table A1 shows the percentage explained by the climate models and SDMs estimated considering
9 two and three sources of variance. The percentages for CY are not shown for summer because EPI
10 could not be calculated for a large number of cases, and the percentages for winter do not include
11 the results from BCM and BCMV. The percentage explained by the GCM-RCM simulations and
12 the SDMs is similar when considering two or three sources of variances. Additionally, the
13 conclusion on which is the most important source of variance is the same for all catchments except
14 for DE and PL in winter. For these two catchments, the percentage explained by the GCM-RCM
15 simulations is approximately 50%.

16 **TABLE A1**

17 **A.2 Comparison of reconstructed and original values**

18 A similar verification approach as the one carried out in Déqué et al. (2007) has also been used. It
19 consists in removing the data for one combination of RCM-GCM and using the matrix
20 reconstruction approach to estimate its values for all SDMs. The reconstructed values are then
21 compared with the original values and also with two other combinations of RCM-GCMs (one using
22 the same RCM and one using the same GCM). This test is applied to two RCM-GCM simulations:
23 RCA-ECHAM5 and HIRHAM-BCM.

24 The reconstructed vector for these combinations is referred to as \mathbf{EPI}_{RG} . In the case of RCA-
25 ECHAM5, \mathbf{EPI}_{RG} is compared with the vectors found for: (i) the original EPI values found for
26 RCA-ECHAM5; (ii) the combination RCA-BCM (\mathbf{EPI}_R in Table A2); (iii) and the combination
27 REMO-ECHAM5 (\mathbf{EPI}_G in Table A2). In the case of HIRHAM-ARPEGE, \mathbf{EPI}_{RG} is compared with
28 the original values, with HIRHAM-ARPEGE (\mathbf{EPI}_R), and RCA-BCM (\mathbf{EPI}_G). Table A2 shows the
29 average of the RMSE obtained for all the catchments, T-yr levels, seasons, and temporal
30 aggregations.

1 Table A2 shows that in the case of RCA-ECHAM5, the difference between the reconstructed and
2 the original values is smaller than the difference between the reconstructed values and the other two
3 RCM-GCM combinations. However, in the case of HIRHAM-BCM, the difference between the
4 reconstructed and the original values is higher than the difference between the reconstructed and the
5 other two RCM-GCM combinations.

6 TABLE A2

7 This results show that in some cases the reconstructed values can differ more from the original
8 values than they differ from other models. Hence, the variances estimated in the variance
9 decomposition approach are likely to be affected by the reconstructed values.

10

- 1 Table 1 - Summary of the main characteristics of the catchments. The column with label “extremes” indicates the season where most
- 2 precipitation extremes occur. The catchments are sorted from north to south, with the most northern catchment in the top row.

Name	River, Country	Area [km ²]	Median altitude [m]	Data used for calculation of catchment precipitation	Mean annual precipitation [mm yr ⁻¹]	Extremes	Observation period
NO2	Nordelva, Norway	207	349	1x1 km grid (Tveito et al., 2005)	2437	Winter	1957 – 2010
NO1	Atna, Norway	463	1204	1x1 km grid (Tveito et al., 2005)	852	Summer	1957 – 2010
DK	Aarhus Å, Denmark	119	65	10x10 km grid (DMI, 2012)	868	Summer	1989 – 2010
LT	Merkys, Lithuania	4416	109	1 station	658	Summer	1961 – 1990
BE	Grote Nete, Belgium	383	32	6 stations	828	Summer	1986 – 2003
DE	Mulde, Germany	6171	414	43 stations	937	Summer	1951 – 2003

CZ2	Upper Metuje, Czech Republic Jizera,	67	588	1x1 km grid (Šercl, 2008)	788	Summer	1980 – 2007
CZ1	Czech Republic	2180	365	10 stations	860	Summer	1951 – 2003
PL	Nysa Kłodzka, Poland	1083	316	2 stations	589	Summer	1965 – 2000
TR	Gocbeylidere, Turkey	609	153	1 station	850	Autumn	1960 – 1990
CY	Yermasoyia, Cyprus	157	575	2 stations	640	Winter	1986 – 1997

1

2

1 Table 2 – Matrix of RCM-GCM combinations used in this study and source of the RCMs.

RCM\GCM	ECHAM5	BCM	HadCM3- Q3	HadCM3- Q16	HadCM3- Q0	ARPEGE	Institute
RM5.1						X	National Centre for Meteorological Research in France
RACMO2	X						Royal Netherlands Meteorological Institute
RCA	X	X	X				Swedish Meteorological and Hydrological Institute
REMO	X						Max Planck Institute for Meteorology
RCA3				X			Community Climate Change Consortium for Ireland
CLM					X		Swiss Federal Institute of Technology
HadRM3Q0					X		UK Met Office
HadRM3Q3			X				UK Met Office
HadRM3Q16				X			UK Met Office
HIRHAM5	X	X				X	Danish Meteorological Institute
RegCM3	X						International Centre for Theoretical Physics

2

3

- 1 Table 3 – Summary of the advantages and disadvantages of each statistical downscaling method. The name of the institution that undertook
 2 the downscaling work in this study is included in the first column. The advantages and/or disadvantages which are specific to the way the
 3 methods are applied in this application are stated.

SD method	Advantages	Disadvantages
Bias correction of mean, BCM (T. G. Masaryk Water Research Institute, Faculty of Environmental Sciences)	Easy to apply and little computer time required. Preserves the sequences of dry/wet days from the RCM. It accounts for different corrections in different time windows.	It only corrects the mean precipitation of the RCM.
Bias correction of mean and variance, BCM (T. G. Masaryk Water Research Institute, Faculty of Environmental Sciences)	(same as bias correction of mean) It allows for distinct corrections between mean and variance.	The non-linear transformation may lead to unexpectedly large precipitation amounts. The autocorrelation from the RCM is not corrected, but it is affected by the bias correction approach.
Bias correction	Easy to apply and little computer time required.	The correction of the upper tail is based on relatively

quantile mapping, BCQM (NVE)	<p>Preserves the sequences of dry/wet days from the RCM.</p> <p>Distinction between corrections in mean and extreme precipitation.</p> <p>The frequency of precipitation is corrected.</p> <p>No theoretical distribution is assumed.</p>	<p>few values (empirical distribution based).</p> <p>In this application, the same correction is applied for all seasons.</p> <p>The autocorrelation from the RCM is not corrected, but it is affected by the bias correction approach.</p>
Expanded downscaling, XDS (U. Potsdam)	<p>Generates realistic weather consistent with large-scale atmospheric patterns.</p> <p>Able to employ full range of predictor variables.</p> <p>It preserves co-variability between the predictands.</p>	<p>High demand for climate model accuracy; systematic biases can cause large errors.</p> <p>Requires large computation time and data preparation.</p> <p>No fully objective way of selecting the predictors.</p>
Change factor of mean, CFM (DHI, DTU)	<p>Easy to apply and little computer time required.</p> <p>It accounts for different changes in different months.</p>	<p>It only accounts for changes in mean precipitation.</p> <p>Does not account for changes in the length of dry/wet spells.</p>
Change factor of mean and variance, CFMV (DHI, DTU)	<p>(same as change factor of mean)</p> <p>Distinction between changes in mean and variance.</p>	<p>Does not account for changes in the length of dry/wet spells.</p> <p>The autocorrelation of precipitation may be disturbed.</p> <p>The non-linear transformation may lead to unexpectedly large precipitation amounts.</p>

Change factor quantile mapping, CFQM (DTU)	(same as change factor of mean) Distinction between changes in mean and extreme precipitation. No theoretical distribution is assumed.	Does not account for changes in the length of dry/wet spells. The changes in the tails are based on relatively few values. The autocorrelation of precipitation may be disturbed.
Change factor quantile perturbation, CFQP (KU Leuven)	(same as change factor quantile mapping) Changes in the frequency of precipitation are accounted for.	The changes in the tails are based on relatively few values. The autocorrelation of precipitation may be disturbed (in this application, this is checked).

1 Table A1 – Percentage of the total variance explained by the GCM-RCM simulations (G+R) and
 2 SDMs (S) considering 2 and 3 sources of variance. The contribution of the GCMs and RCMs is
 3 shown in brackets.

	Nr. sources	Winter		Summer	
		G+R	S	G+R	S
NO2	2	68	32	51	49
	3	69 (29+40)	31	52 (14+38)	48
NO1	2	51	49	60	40
	3	51 (13+38)	49	61 (13+48)	39
DK	2	60	40	65	35
	3	62 (22+40)	38	67 (26+41)	33
LT	2	59	41	60	40
	3	57 (20+37)	43	57 (10+47)	43
BE	2	69	31	51	49
	3	71 (30+41)	29	52 (15+37)	48
DE	2	49	51	62	38
	3	51 (18+33)	49	61 (16+45)	39
CZ2	2	54	46	61	39
	3	55 (15+41)	45	57 (14+43)	43
CZ1	2	60	40	64	36
	3	58 (24+34)	42	59 (19+40)	41
PL	2	51	49	55	45
	3	48 (21+28)	52	50 (19+30)	50
TR	2	57	43	46	54
	3	55 (19+35)	45	42 (19+23)	58
CY	2	55	45		
	3	55 (21+34)	45		

4

1 Table A2 – Average RMSE from the comparison of the reconstructed and original values and the
2 comparison with other combinations of GCM-RCM

RCM\GCM	Original	EPI_R	EPI_G
RCA-ECHAM5	0.47	0.60	0.61
HIRHAM-BCM	2.49	1.46	2.45

3

4

1 Figure 1 - Location of the eleven catchments studied.

2

3 Figure 2 – EPI estimated from the comparison of the downscaled time series for control and future
4 period for 1 yr (light grey boxes) and 5 yr levels (dark grey boxes). The boxes indicate the 25, 50
5 and 75th percentiles and the whiskers the 5 and 95th percentiles. The circles show the median of all
6 the values of EPI estimated from the comparison of the RCM outputs for the control and future
7 periods. All the results are for a temporal aggregation of 1 day.

8

9 Figure 3 – In the top row, total variance decomposed in variance from GCMs, RCMs, SDMs and all
10 the interaction terms (darkest to lighter grey colours). In the bottom row, percentage of the total
11 variance explained by GCMs, RCMs, and SDMs (darkest to lighter grey colours). All the results are
12 shown for 1 and 5 yr levels in the left and right column of each catchment, respectively. All the
13 results are for a temporal aggregation of 1 day.

14

15 Figure 4 – EPI for each SDM for NO₂, DE, and TR for winter (top) and summer (bottom). The
16 markers indicate the median and the lines represent the range covered by the 25th and 75th
17 percentiles. All results are for the 1 yr level and temporal aggregation of 1 day. Note the different
18 scales used in the y-axis for winter and for summer.

19

20 Figure 5 - EPI for each SDM for NO₂, DE, and TR for winter (top) and summer (bottom). The
21 markers indicate the median and the lines represent the range covered by the 25th and 75th
22 percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal
23 aggregation. The same symbols are used for the different downscaling methods as in Fig. 4. Note
24 the different scales used in the y-axis for winter and for summer.

25

1 Figure 6 – EPI estimated from the comparison of the observations and the downscaled time series
2 by all BC methods for the control period for 1 yr (light grey boxes) and 5 yr levels (dark grey
3 boxes). The boxes indicate the 25, 50 and 75th percentiles and the whiskers the 5 and 95th
4 percentiles. The circles show the median of all the values of EPI estimated from the comparison of
5 the observations and the uncorrected RCM outputs for the control period. All the results are for a
6 temporal aggregation of 1 day.

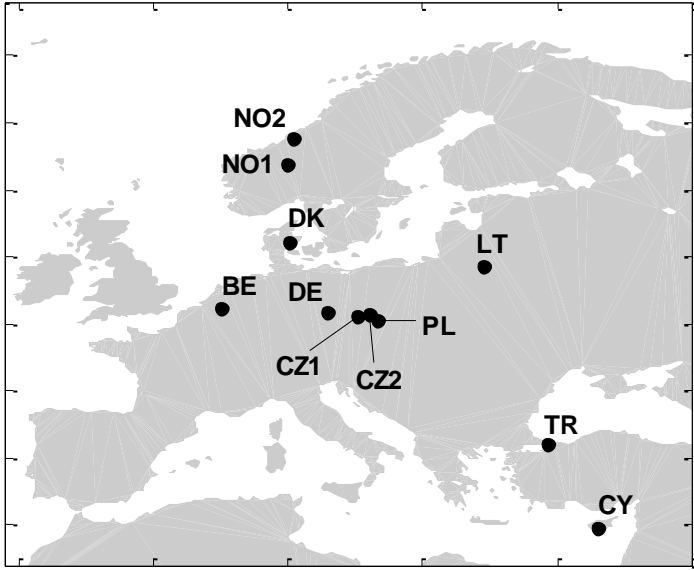
7

8 Figure 7 - EPI for each BC method for NO₂, DE, and TR for winter (top) and summer (bottom).
9 The markers indicate the median and the lines represent the range covered by the 25th and 75th
10 percentiles. All the results are for the 1 yr level and temporal aggregation of 1 day. Note the
11 different scales used in the y-axis for winter and for summer.

12

13 Figure 8 - EPI for each BC method for NO₂, DE, and TR for winter (top) and summer (bottom).
14 The markers indicate the median and the lines represent the range covered by the 25th and 75th
15 percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal
16 aggregation. All the results are for 1 yr threshold. The same symbols are used for the different
17 downscaling methods as in Fig. 7. Note the different scales used in the y-axis for winter and for
18 summer.

19

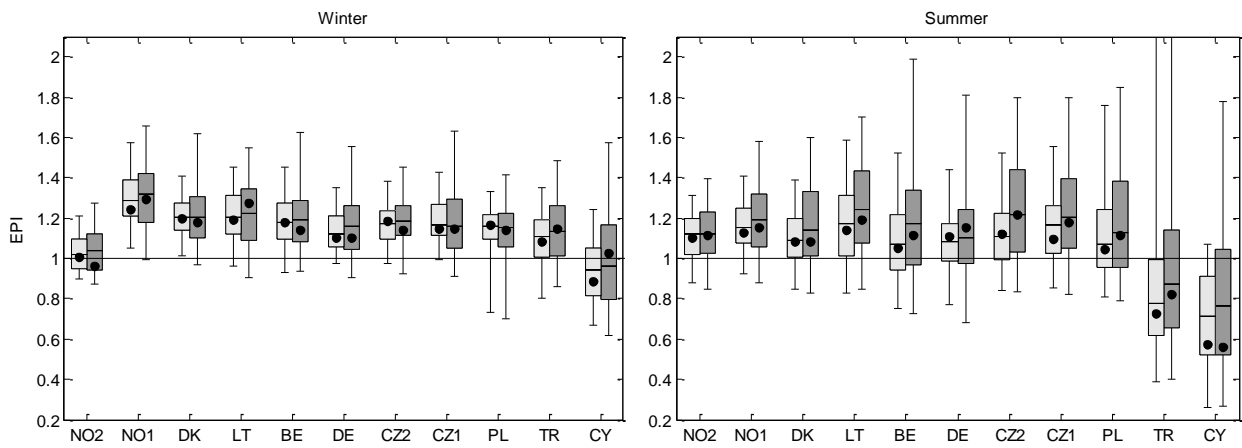


1

2 Figure 1 - Location of the eleven catchments studied.

3

4

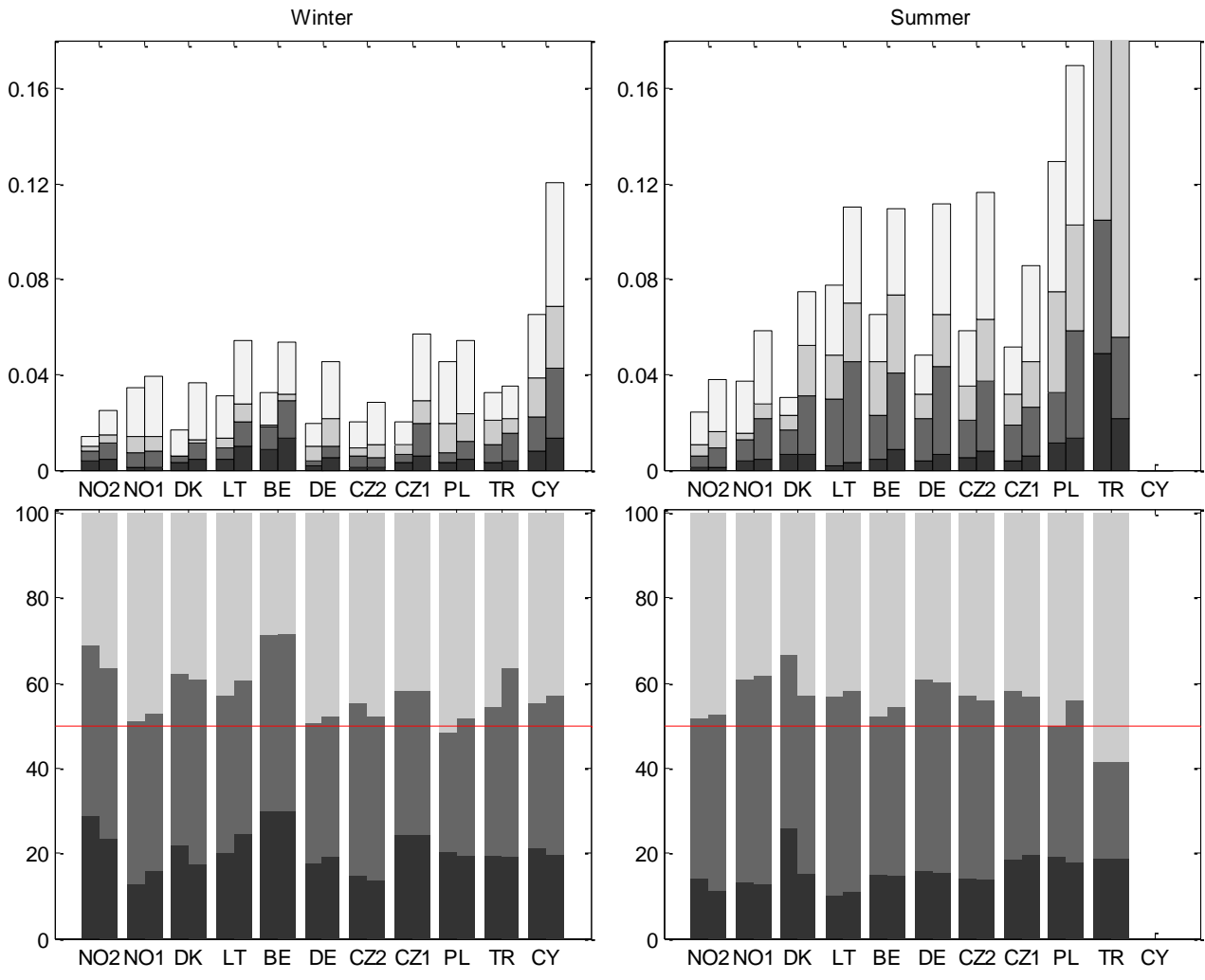


1

2 Figure 2 – EPI estimated from the comparison of the downscaled time series for control and future
 3 period for 1 yr (light grey boxes) and 5 yr levels (dark grey boxes). The boxes indicate the 25, 50
 4 and 75th percentiles and the whiskers the 5 and 95th percentiles. The circles show the median of all
 5 the values of EPI estimated from the comparison of the RCM outputs for the control and future
 6 periods. All the results are for a temporal aggregation of 1 day.

7

8

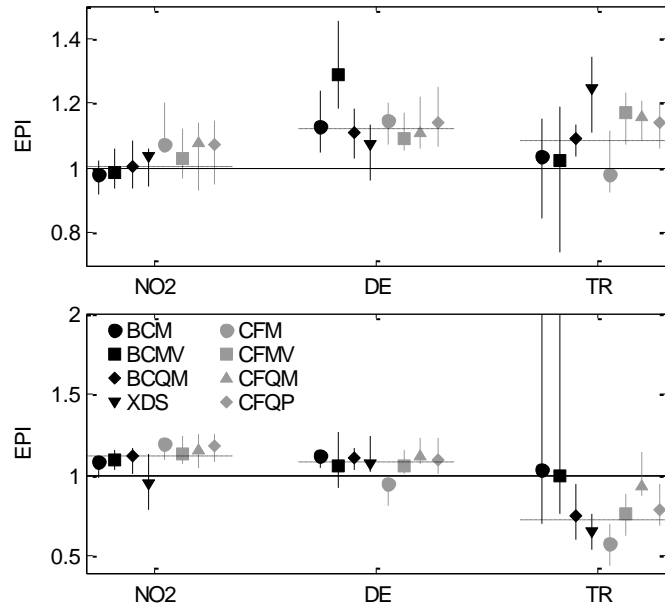


1

2 Figure 3 – In the top row, total variance decomposed in variance from GCMs, RCMs, SDMs and all
 3 the interaction terms (darkest to lighter grey colours). In the bottom row, percentage of the total
 4 variance explained by GCMs, RCMs, and SDMs (darkest to lighter grey colours). All the results are
 5 shown for 1 and 5 yr levels in the left and right column of each catchment, respectively. All the
 6 results are for a temporal aggregation of 1 day.

7

8

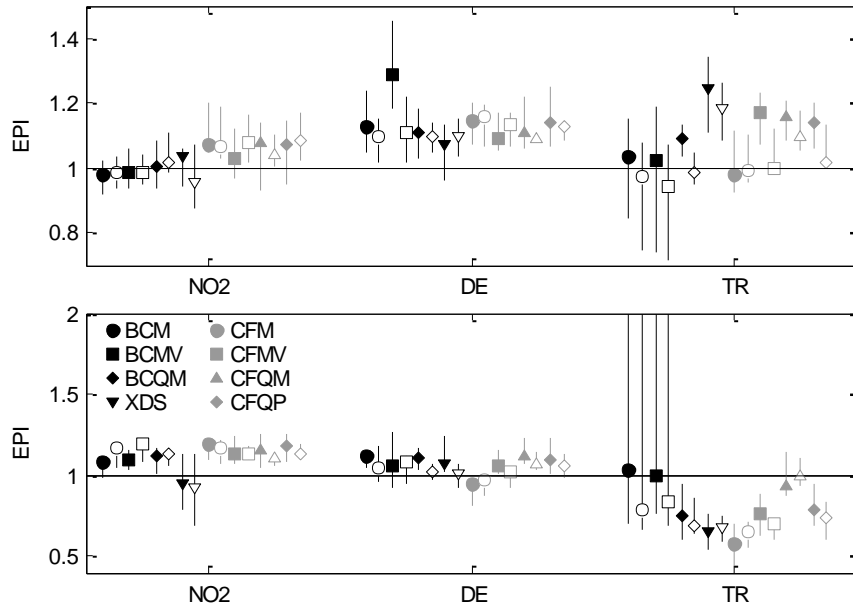


1

2 Figure 4 – EPI for each SDM for NO2, DE, and TR for winter (top) and summer (bottom). The
 3 markers indicate the median and the lines represent the range covered by the 25th and 75th
 4 percentiles. All results are for the 1 yr level and temporal aggregation of 1 day. Note the different
 5 scales used in the y-axis for winter and for summer.

6

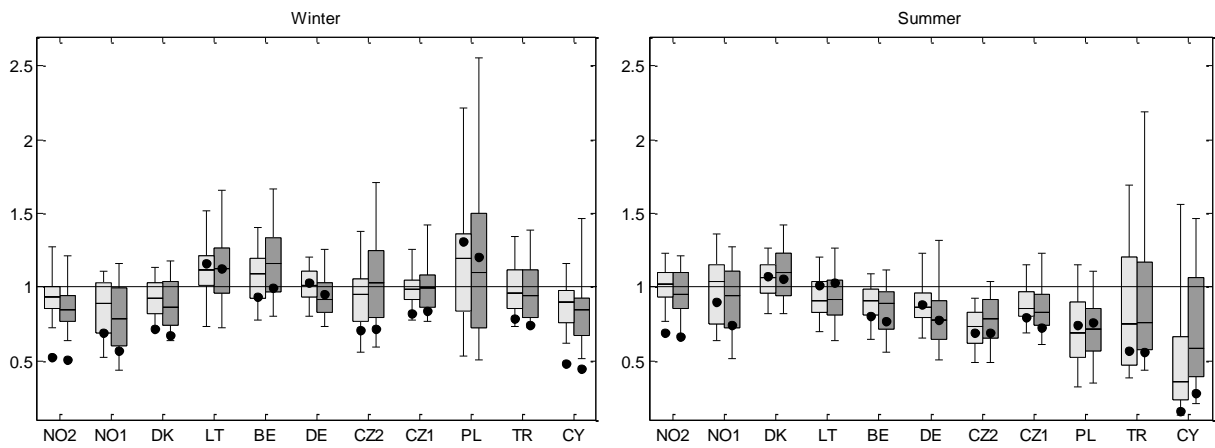
7



1

2 Figure 5 - EPI for each SDM for NO2, DE, and TR for winter (top) and summer (bottom). The
 3 markers indicate the median and the lines represent the range covered by the 25th and 75th
 4 percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal
 5 aggregation. The same symbols are used for the different downscaling methods as in Fig. 4. Note
 6 the different scales used in the y-axis for winter and for summer.

7

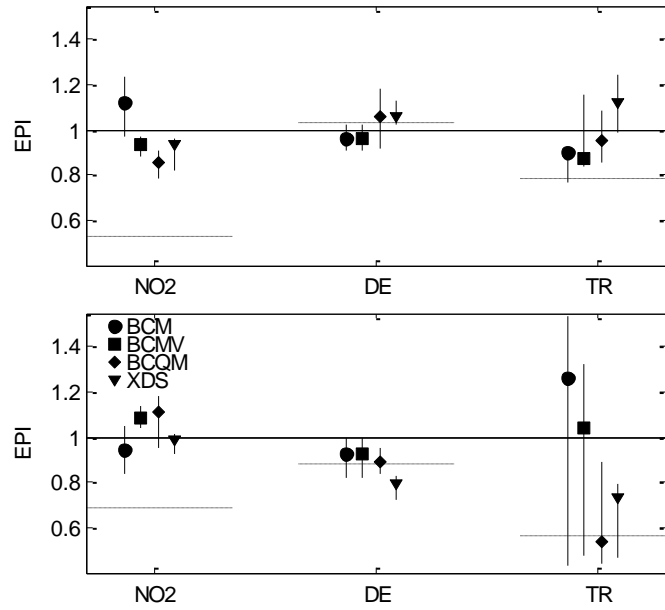


1

2 Figure 6 – EPI estimated from the comparison of the observations and the downscaled time series
 3 by all BC methods for the control period for 1 yr (light grey boxes) and 5 yr levels (dark grey
 4 boxes). The boxes indicate the 25, 50 and 75th percentiles and the whiskers the 5 and 95th
 5 percentiles. The circles show the median of all the values of EPI estimated from the comparison of
 6 the observations and the uncorrected RCM outputs for the control period. All the results are for a
 7 temporal aggregation of 1 day.

8

9



1

2 Figure 7 - EPI for each BC method for NO2, DE, and TR for winter (top) and summer (bottom).

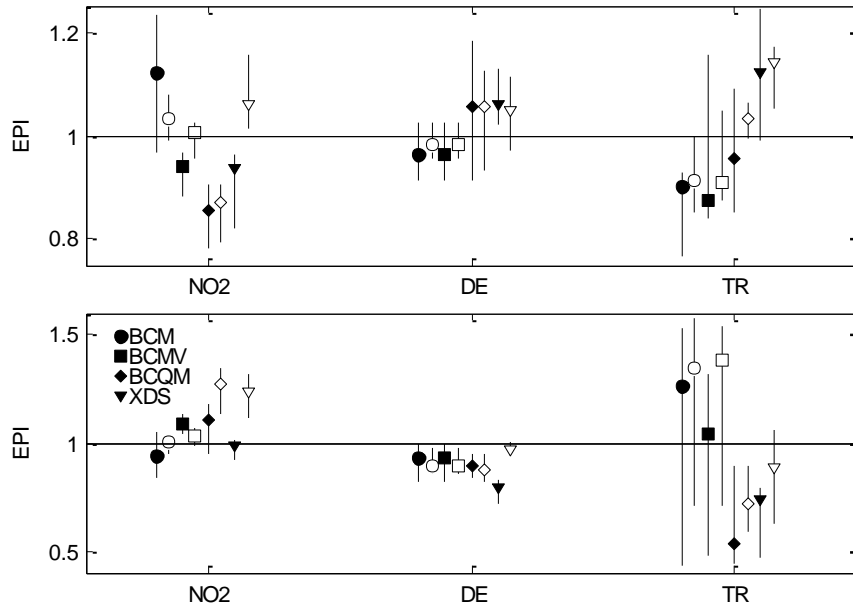
3 The markers indicate the median and the lines represent the range covered by the 25th and 75th

4 percentiles. All the results are for the 1 yr level and temporal aggregation of 1 day. Note the

5 different scales used in the y-axis for winter and for summer.

6

7



1

2 Figure 8 - EPI for each BC method for NO₂, DE, and TR for winter (top) and summer (bottom).
 3 The markers indicate the median and the lines represent the range covered by the 25th and 75th
 4 percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal
 5 aggregation. All the results are for 1 yr threshold. The same symbols are used for the different
 6 downscaling methods as in Fig. 7. Note the different scales used in the y-axis for winter and for
 7 summer.

8