Responses referee #1 on the article HESS-2011-117 entitled
"Improving pan-European hydrological simulation of extreme
events through statistical bias correction of RCM-driven climate
simulations"
by R. Rojas et al.

June 24, 2011

We would first like to thank the constructive comments on our manuscript raised by the reviewers. In the next sections, we provide a detailed list with the responses to the major remarks pointed by the reviewers hoping to clarify remaining issues. We proceed by listing the reviewers' comments (**in bold text**) and the corresponding reply.

Anonymous referee #1

General

C1. The paper describes the application of a previously developed bias correction method at the European scale using a new dataset. The paper is generally well-written and clearly structured, and the results are clearly presented and discussed. However, the results that are presented are not very surprising, because Piani et al, (2010) already showed that the method worked well at the global scale. Moreover, the current study does not use a validation period, as, for example, Piani et al., (2010) did (see below). This would assess the performance of the method much more thoroughly. Another addition that also would make the paper also much more interesting would be to compare the performance of multiple bias correction methods. As far as I know that has not been done very often (apart from maybe Themessl et al, (2010) for an Alpine region), especially not at the European scale. I would like to see that the authors include a validation of the method, i.e., application to part of the dataset for which it was not calibrated, at least for some sub-areas.

A1. In Piani et al. (2010a) and Piani et al. (2010b), respectively, the bias correction (BC) method was applied at Pan-European and global scale to correct precipitation and temperature fields only. In our work we are building upon these two studies to assess the value of the BC in the simulation of hydrological extreme events at pan-European scale. Given the scale, the spatial and time resolutions, and the setting-up of the hydrological model employed, our study shows quite a few novel elements compared to previous literature. At the same time, comparing alternative BC methods is beyond the scope of this work, as we are interested in assessing the hydrological impacts of the BC method on extreme events, and this method proves to work very good for our purposes. As mentioned by the reviewer this has been done in the past by several authors at the basin-scale, although not at pan-European scale.

One of the main concerns of the reviewer is the non-existing validation period to assess the BC method. In a parallel work, one of the co-authors of this study has assessed the performance of the BC method for 11 RCMs contained in the ENSEMBLES project (Dosio and Paruolo, 2011). In his study, the same calibration period (1961–1990) as the one employed in our work and a validation period of 10 years (1991–2000) are defined to assess the BC method. One of the RCMs analyzed corresponds to our study case (bottom-right panel of Figure 1). Therefore, we believe that it is unnecessary to repeat the work done by Dosio and Paruolo (2011). As an example of the validation results we include the precipitation bias for summer (JJA) for the validation period 1991–2000, obtained with transfer functions calculated in the period 1961–1990 for 11 RCMs from the ENSEMBLES project. Figure 1 shows that the BC method is performing acceptably in the validation period with exception of a few localized areas.

Specific remarks

SR1. Page 3889, line 20: (e.g. 0.5 and 0.25 regular...). E.g. should be i.e. I assume as there are probably not more than 2 resolutions. Please specify here which resolution you used. Furthermore, the E-OBS dataset is being repeatedly referred to as high-resolution (also in the Introduction). 25km, however, does not exactly seem high-resolution, given that e.g., Dankers et al., (2007) already applied HIRHAM and LISFLOOD at 12 km for entire Europe.

A1. Corrected (see revised manuscript). We have indicated the use of 0.22 rotated pole grid. Dankers et al. (2007) used data from HIRHAM (ca. 12km) to simulate flood hazard in the upper Danube Basin only. Dankers and Feyen (2008) used data from HIRHAM (ca. 12km) to simulate flood hazard at Pan-European scale using a 5km grid resolution for LISFLOOD. From this it seems the referee is mixing
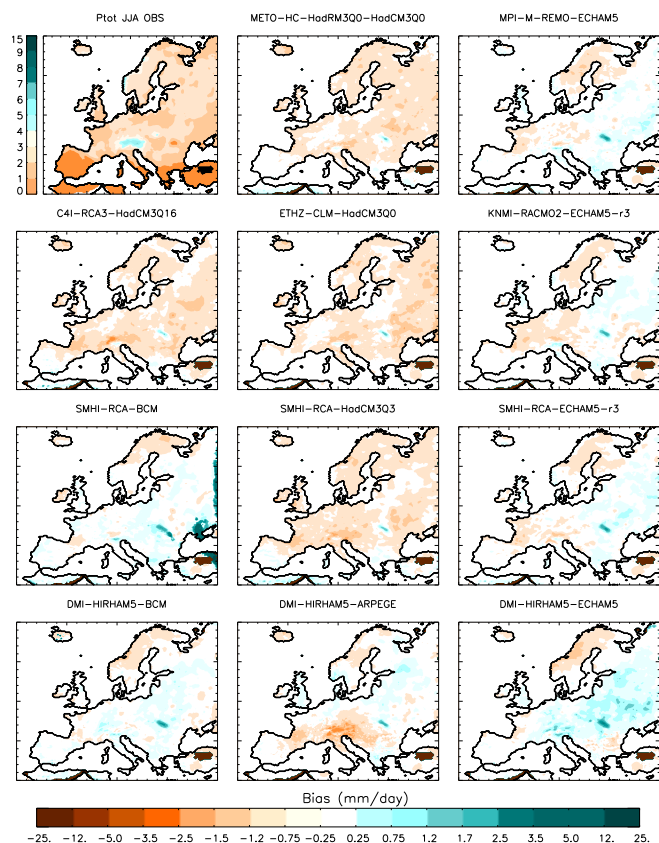
Figure 1: Summer observed mean daily precipitation (E-OBS dataset) for the validation period 1991-2000 (left-top panel) and bias as modeled by the ENSEMBLES RCMs (Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: evaluation on the present climate, Journal of Geophysical Research, doi:10.1029/2011JD015934, 2011. Copyright 2011 American Geophysical Union. Reproduced by permission of the American Geophysical Union).

observation with simulated data sets. The climate data employed in this work is referred to as "high-resolution" given the facts that preliminary ensembles of climate models were predominately run at 50km (e.g. PRUDENCE project). In addition, it seems correct to refer to the E-OBS as high-resolution as it is the data set with the highest resolution for Europe, and is often referred in this terms (see title of the original article: Haylock, M., Hofstra, N., Klein, A., Klok, E., Jones, P., and New, M.: A European daily **high-resolution** gridded data set of surface temperature and precipitation for 19502006, Geophysical Research Letters, 113, doi:10.1029/2008JD010201, 2008).

SR2. Page 3894, lines 13-18: As mentioned above, my main concern is the lack of a real validation period. The problem is that the construction period and control period are the same. Because the transfer functions are derived for 1961-1990 it is not very surprising the results are good for the same period. A proper validation would mean, for example, deriving TFs for 1960-1969 and validating them on 1990-1999 (as Piani et al., 2010 did). More generally, one of the methods outlined in Klemes (1986) should be used.

A2. See A1 in section General. Corrected (see revised manuscript). The reader is referred to the work of Dosio and Paruolo (2011).

SR3. Page 3894, lines 18-22: Here, and also at various other places in the manuscript, it is stressed that the comparison between the control and future period are subject to the stationarity assumption. I appreciate that there are no real alternatives for this assumption, but it should be discussed that it is most probably not valid, as outlined by Christensen et al. (2008; reference already in manuscript), who found that biases depend on temperature and precipitation values and can grow under climate change conditions.

A3. Corrected (see revised manuscript). The reference was added.

SR4. Page 3895, lines 16-20: The information provided about how LISFLOOD was calibrated is quite brief. First of all, 4 years is a short time to calibrate a model for such a long-term study. The chance that any hydrological extremes are included in such a period is small, although this is important as the parameters are assumed to be valid under changed climate conditions. Is the forcing data to calibrate LISFLOOD (the MARS database) purely based on observations or are also models involved? Whats its resolution? Is there any information about how well MARS corresponds to E-OBS? Possibly the model calibration could be based on biased forcing data (wrt. E-OBS). Would it not have been possible to calibrate the model using E-OBS?

A4. Corrected (see revised manuscript). We agree with the fact that four years is a short time to calibrate a hydrological model for a long-term study. The selection of four years responded to a trade-off between computational time and the use of reliable and recent available information on discharges. Even if the calibration period is restricted to 4 yr, we must stress that LISFLOOD was calibrated in the framework of the European Floods Alert System (EFAS), thus, having a particular interest in correctly reproducing the timing and magnitude of flooding events.

We must emphasize that the calibration of LISFLOOD at pan-European scale is an enormous task which is constantly updated in different calibration rounds. In our work, we used the fourth calibration round of LISFLOOD. This calibration run employed mainly forcing data coming from the MARS database locally complemented with some data from the SYNOP database. Point measurements for precipitation, average, maximum and minimum temperatures, wind speed, radiative forcing, and dewpoint temperatures were interpolated on a grid of 5×5km to obtain the input data forcing LISFLOOD during the calibration period. The interpolation method was the inverse weighted distance. Discharge data spanning at least 4 yr between 1995-2002 was selected for 258 basins and the main algorithm used to calibrate LISFLOOD was the SCE-UA (Duan et al., 1992). More details about the calibration, e.g. parameters to be calibrated, location of calibration stations, and the calibration algorithm implemented, can be found in the references included in the manuscript.

Unfortunately, there is no information available nor previous works on how well the MARS database corresponds to the newly developed E-OBS data set. Regarding the calibration of LISFLOOD using the E-OBS data set as forcing data this is not feasible for the time being. To calibrate LISFLOOD forced by the E-OBS data set key information is missing. From the E-OBS we only have four fields (daily avg, max and min temperatures and precipitation), thus lacking relevant fields for the calculation of the evapotranspiration demands used to force LISFLOOD (e.g., radiative forcings, dewpoint temperatures, wind speed, albedo). Therefore, a new calibration of LISFLOOD forced by E-OBS is not feasible. We must add that currently the fifth calibration round of LISFLOOD is being implemented within a three-year research task. The new calibration round will combine more databases (MARS, SYNOP, ECA&D) to obtain validated input forcing data, more stations with validated discharge data (489 instead of 258), and it will improve on the calibration algorithm and the assessment of the calibration results.

SR5. Page 3900, Figure 3: Although the wet day frequency is improved, the bias correction overshoots the observations, causing an underestimation throughout Europe. This is not commented on. Why is that? Similarly, at Page 3901, line 12 the authors note that 3, 5, and 7 day totals are underestimated in the after correction whereas they were overestimated before. Is there an explanation for that?

A5. We fully agree with the reviewer on the fact that the wet day frequency for bias corrected climate data slightly underestimates observations. This underestimation, however, is not as severe as presented in Figure 3 of the manuscript. In Figure 2 we include the wet days frequency for uncorrected (panel a), ENS-OBS (panel b) and bias corrected (panel c) climate data. An excellent correspondence for the wet days frequency between ENS-OBS and bias corrected climate data is observed with some minor and localized discrepancies in northern Scandinavia and Eastern Europe. Spatially, the observed pattern is fully preserved.

Unfortunately, in Figure 3 of the manuscript we have selected a reclassification to emphasize the areas of overestimation (panel a in Figure 3 of the manuscript) which gave the false impression of a strong underestimation of the wet days frequency given the non-symmetric legend of the figure. As observed from the summary statistics included in Figure 3 of the manuscript (bottom of panel b), the annual average underestimation for the whole 30-yr period is about 7 days, which is rather moderate compared
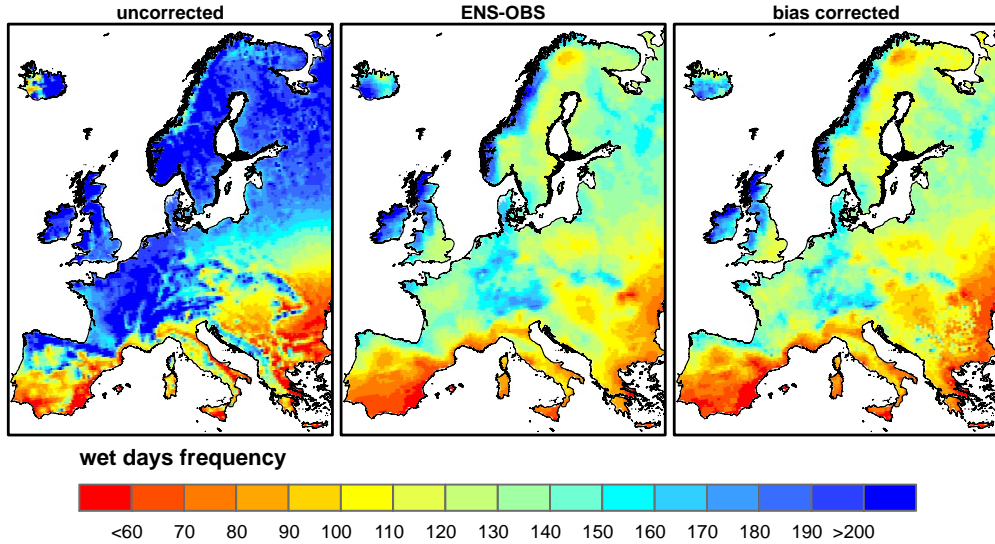
Figure 2: Wet days frequency for a) uncorrected climate data b) ENS-OBS data set and c) bias corrected climate data for the control period 1961–1990.
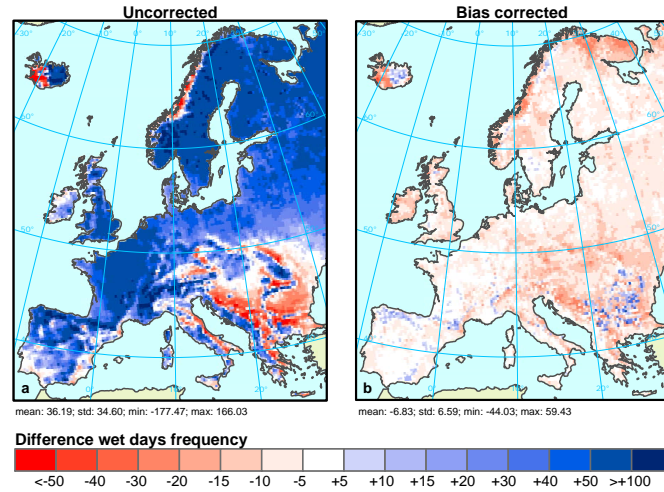

Figure 3: Difference of wet days frequency between a) uncorrected, b) bias corrected precipitation and E-OBS data set for the control period 1961–1990.

to the strong overestimation observed for the uncorrected climate data (36 days). As an example, Figure 3 of this review shows the same figure included in the manuscript with a more symmetric legend emphasizing over- and underestimation at the same time.

From an ensemble analysis, Dosio and Paruolo (2011) suggested that low-end percentiles of bias corrected precipitation are subject to large uncertainties due mainly to the choice of the transfer function (linear vs. exponential) used to perform the bias correction. At the same time, they found a systematic underestimation of the small values of bias corrected precipitation compared to the observed pdf obtained from the E-OBS data set (see Figure 14 in Dosio and Paruolo 2011). Additionally, and as explained in the manuscript, the exponential transfer function employs the dry day correction ($x_0 = -a/b$), defined as the value below which bias corrected precipitation is set to zero. x_0 is obtained directly from the optimized parameters (a, the additive correction factor and b, the multiplicative correction factor of the linear asymptote) to compensate for the number of dry days in the observed precipitation. As highlighted by Piani et al. (2010b), $x_0 > 0$, since observed precipitation shows many more dry days compared to the simulated precipitation from RCMs. This is explained by the systematic bias of RCMs to simulate too many days with too little precipitation (drizzle), thus, underestimating the number of dry days. As a consequence, the underestimation of the wet days frequency after bias correction is likely explained by: first, the systematic underestimation of the low-end percentiles for the bias corrected precipitation

4

observed by Dosio and Paruolo (2011) which increases the probability for a given day to be considered as a "dry day" for a constant x_0, and second, by the potential overestimation of the number of dry days given that x_0 is obtained from the optimized parameters a and b solely, which may differ from the actual number of dry days in the observed precipitations.

To avoid ambiguities, we have rephrased the corresponding sections emphasizing the reasons behind the underestimation of the wet day frequency and 5-days average maximum precipitation.

SR6. Page 3903/3904, Figure 8: The effect of bias correction on evapotranspiration is quite large. This is surprising considering that ET was calculated using Penman-Monteith, which is mainly radiation based and temperature only enters the equation indirectly through the vapor pressure deficit. As is explained in the manuscript, only temperature is corrected and all other variables remain constant. Do these changes in ET really come from the temperature correction only? In that case the spatial pattern in Figure 8 should be similar to the temperature patterns in Figure 6 and 7, whereas the latter seem much smoother. Is something else than temperature changing as well? Please explain.

A6. We stated in the manuscript that precipitation together with average, maximum and minimum temperatures are bias corrected. We employed the Penman-Monteith equation to calculate potential evapotranspiration demands as implemented in LISVAP (van der knijff, 2008). Average temperature is used to calculate the latent heat of vaporization and saturated vapor pressure. Tmax and Tmin are used to calculate the wind coefficient of the wind function to correct the evaporative demand of the atmosphere. Therefore, strictly speaking, average, maximum and minimum temperatures influenced ET calculations. Since uncorrected average temperatures are overestimating observed temperatures we should expect larger areas with a higher ratio ET_{unced}/ET_{bced}.

SR7. Page 3905, Figure 10: It is not entirely clear to me what is plotted. Are these 30 year averages/maxima for all basins (554 points), or annual averages/maxima (30*554 points)? Also, what do you mean by model efficiency, is that Nash-Sutcliffe coefficient? If thats the case, and the plot refers to 30 year averages, it does not make much sense to calculate a Nash-Sutcliffe because that indicates the efficiency with respect to an average. Or is it the model efficiency averaged over all basins? In that case, a value of 0.99 is not really credible, even when the temporal resolution is only annual.

A7. Corrected (see revised manuscript). Series plotted in Figure 10 of the manuscript correspond to average discharge and average annual maximum discharge for each of the 554 stations. We used model efficiency as defined by the Nash-Sutcliffe criterion. In principle, we are comparing (30 years) average discharges and (30 years) average annual maxima discharges, thus, we are measuring the efficiency with respect to those average predictions. This is **not** the model efficiency averaged over all basins.

SR8. Page 3906/Figure 11: It is not clear what the black crosses are exactly. They are said to be observations but in cases go up to a return period of about 100 yr, whereas about 30 years were available (page 3896)? If they are based on a Gumbel-fit, why are they not shown as lines? If much more than 30 years of observations is available for these basins, please mention that. Also, the numbers in Table 3 do not seem to be completely consistent with the crosses in Figure 11 for example, the 100-yr discharge for the Daugava (misspelled as Daugana in Table 3 by the way) is about 6000 in Figure 11, and not even 4000 in Table 3.

A8. Black crosses show the return levels obtained from the empirical plotting position for the 30 annual maxima discharges for each gauging station depicted in Figure 11. Return levels are obtained as $rl = -1/log(f)$ where f is the plotting position of the sorted (in increasing order) series of annual maximum values given by $i/n + 1$, where i is the position in the ordered series and n the total number of data values. We employed plotting positions to examine the intrinsic variability of the annual maxima. They cannot go up to 100 yr, they actually reach a maximum of ca. 70 yr given that we are working with 30 annual maxima. Indeed, as written in Table 3 of the manuscript values under the "observed" heading correspond to fitted Q_{100} using a Gumbel distribution obtained from the observations, this to allow a fair comparison in the table. Therefore, in Figure 11 we have return levels obtained from empirical plotting positions from 30 observed annual maximum discharges for each of the 20 stations, whereas in Table 3 we have the Q_{100} value obtained from the fitting of a Gumbel distribution to the observations. To avoid ambiguity, text explaining figure 11 and caption of Table 3 have been corrected.

SR9. Page 3906/Table 3: how is the percentage reduction calculated? I did not manage to get exactly the same numbers from the confidence intervals in the table. Sometimes close, but sometimes way off, e.g. for the Danube the reduction seems much more than 2.5%, and the Nemunas much less than 28.5%. If the Gumbel-fits themselves (ie not the confidence intervals) are used, please show those as well, and check for switched or miscalculated numbers.

A9. First the length of the confidence interval is calculated for both corrected and uncorrected as $l_i = up95_i - low95_i, i = 1, \ldots, 20$. Then we calculate $l_{unc} - l_{bced}/l_{unc}$ as the percentage reduction. We have verified the numbers in Table 3, and we have corrected two typing errors.

SR10. Page 3908, lines 26-28: I do not agree with the notion that the future recurrence interval does only depend on the discharge magnitude change between control and future period (also mentioned in the Conclusions). The point is that the discharge magnitude is the same (and thus does not change) when you compare two recurrence intervals based on two Gumbel-fits. To obtain those Gumbel-fits, you do need time series of (annual maximum) discharge magnitudes, both for the control and future period. Whereas the authors seem to use the correct method for this comparison, it is explained rather confusingly. Please reformulate. However, because the time series are only 30-years, the 100-year discharges are based on very uncertain extrapolations as the authors correctly noted. It would be worthwhile to consider also shorter intervals (30 years or less) that are less prone to the uncertainty of the fit.

A10. Corrected (see revised manuscript). As pointed out by the reviewer, we used indeed two time series of 30-yr annual maxima discharges (control 1961–1990 and future 2071-2100). For each of this time series a Gumbel distribution was fitted. Discharges associated to a 100-yr flood event observed in the control period were obtained for both uncorrected and bias corrected simulations. To obtain the deviations from the control 100-yr flood event we plugged in the discharges of 100-yr flood event of the control period into the Gumbel distribution fitted for the future time slice (2071–2100). Therefore, deviations from a value of 100 would indicate whether the control discharge associated to a 100-yr event would be more frequent or not.

SR11. Page 3908: The grammar here gets a bit sloppy, e.g., ..bias-corrected driven simulations should be something like simulations driven by bias-corrected forcing data. Similar formulations occur several times, and a few more examples (not all) from other sections are noted in the Technical comments. Please check for grammar mistakes.

A11. Corrected (see revised manuscript).

Technical comments

TC1. Page 3884, line 6: Predictand is not the correct term here, as no prediction is being carried out, and, more importantly, the observations are the target and not the variable that is to be predicted.

A1. Corrected (see revised manuscript).

TC2. Page 3887, lines 18-27: This is a very long and confusing sentence. Please split into segments (e.g. for each method) or indicate the methods with i), ii), iii) etc.

A2. Corrected (see revised manuscript).

TC3. Page 3902, line 18: is A function of...

A3. Corrected (see revised manuscript).

TC4. Page 3892, line 8: Xcor and Xsim were explained before and can be removed here

A4. Corrected (see revised manuscript).

TC5. Page 3893, line 4: functions are obtainED for ...

A5. Corrected (see revised manuscript).

TC6. Page 3893, line 8: weighted linear interpolation ... weighted based on what?

A6. Corrected (see revised manuscript).

TC7. Page 3894, line 12: 5 and 0.2 should be reversed.

A7. Corrected (see revised manuscript).

TC8. Page 3895, line 24: remove on in discuss on the calibration...

A8. Corrected (see revised manuscript).

TC9. Page 3896, line 9/10: series of daily discharge maps for each river pixel...this is a confusing formulation. I suppose you have time series for all river pixels?

A9. Corrected (see revised manuscript).

TC10. Page 3902, line 18: Tmin and Tmax were corrected indirectly, but the range was corrected for directly. Hence the improved diurnal range is not very surprising.

A10. Deleted (see revised manuscript).

TC11. Page 3906, line 11: ...explain why () the annual maxima FOR the greater part of the stations FALL below...

A11. Explanation is mentioned in the text immediately above.

TC12. Page 3907/Figure 12: The comparison of corrected and uncorrected in text and Figure are opposite I.e, the figure shows uncorrected/corrected (although this is not indicated), whereas the text mentions the corrected values being higher north of the Carpathians. It would be logical to show corrected/uncorrected in Figure 12, in order to present it such that the ratio north of the Carpathians is higher as well.

A12. Figure and explanation in the text are correct. Caption of the figure and text clearly state ratio between uncorrected and bias-corrected driven simulations. As figure 12 is showing unced/bced we could expect, therefore, a reduced ratio north of the Carpathians where bced > unc.

TC13. Page 3916, line 12: The journal for this reference should be JGR, not GRL

A13. Corrected (see revised manuscript).

Bibliography

Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: evaluation on the present cimate, Journal of Geophysical Research, in press, doi:10.1029/2011JD015934, 2011.

Piani, C., Haerter, J., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theoretical and Applied Climatology, 99, 187–192, doi:10.1007/s00704-009-0134-9, 2010a.

Piani, C., Weedon, G., Best, M., Gomes, S., Viterbo, P., Hagemann, S., and Haerter, J.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, Journal of Hydrology, 395, 199–215, doi:10.1016/j.jhydrol.2010.10.024, 2010b.

van der knijff, J.: LISVAP, Evaporation pre-processor for the LISFLOOD water balance and flood simulation model. Revised user manual, Tech. Rep. EUR22639 EN/2, Joint Research Centre, European Commission, 2008.