Responses referee #3 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 29, 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 #3

General comments

C1. The publication is well written, though lengthy at times and repetitive in some places. It is an interesting piece of work with extensive diagrams and analysis, though in some places long winded . It is suggested that the paper might be split into two papers, on to look at the bias correction of the data sets and discuss the various aspects, the second on the use of this data set in conjunction with LISFLOOD and more importantly the effect of climate change on extreme events as the paper suggests. Some of the diagrams are very small and difficult to read, the space on the pages is not used effectively, the axis on the return period plots not used efficiently to display the spread of the data. The previous reviewers have commented on a number of issues in particular in terms of technical corrections and these shall not be repeated here any further, I concur in particular with comments made on validation of the hydrological model and concerns about the discrepancies of the percentages (Table 3).

A1. We thank the referee for the positive remarks on our work. As mentioned in the responses to the different referees' comments, we have taken into consideration all the suggestions to reduce and summarize the text in the article. We expect to meet a balance between the requirements of more details in given sections of the article and summarizing unwanted details.

Recently, Dosio and Paruolo (2011) submitted an in-depth analysis of the same bias correction method for an ensemble of 11 climate models from the ENSEMBLES project. In their work, they analyzed the bias correction method and various aspects as performance, validation, pdfs of bias corrected variables, and proper ways to express relationships among univariate bias corrected fields. We build upon their work by assessing the impact of the BC on the simulation of hydrological extreme events, with particular emphasis on floods. Even if Dosio and Paruolo (2011) cover aspects that we did not include (e.g. validation and assessment of the performance of the BC method), we report specific results of the BC method relevant for the simulations of floods (e.g. wet day frequency, seasonal 5-days precipitation, seasonal 99^{th} percentile, T_{max}, T_{min}). However, these results may not sustain a publication by themselves. We therefore believe that it is not a good strategy to split up the article into two papers.

Regarding the validation of the hydrological model LISFLOOD, we assume the comment is pointing to the fact that LISFLOOD was not run using as forcing data observations from the E-OBS data set. This was already discussed in our reply to referee #2 and for clarity we answer here. To perform a LISFLOOD run forced only by observations (in order to validate the hydrological simulations) is not possible as relevant fields to calculate the evapotranspiration (radiative forcings, dewpoint temperatures, wind speed and albedo) are missing. In addition, it was suggested the possibility of combining different data sets (e.g. E-OBS, CRU, WATCH). We believe that by doing that the task of identifying (and possibly disentangle) the discrepancies between observed and simulated extreme discharges is even harder.

About the discrepancies in Table 3, we have verified the numbers and corrected two typing mistakes.

Specific comments

SC1. The authors have decided to use a Gumbel Distribution instead of the GEV distribution or a combinations of both. They argue that this results in a reduction in terms of uncertainty. Using the GEV simply adds another degree of freedom to the fitting of the distribution, which is not necessarily an increase in uncertainty. Even though that Dankers and Feyen (2008) showed that neither of the GEV or Gumbel function is preferable over the other, this doesnt justify using only one.

A1. We partially agree with this comment. Given that GEV contains an extra parameter (ξ controlling the shape of the tail), return levels associated to higher return periods are more sensitive compared to the same return levels obtained with the Gumbel distribution. In that sense, estimations from the GEV are more uncertain compared to Gumbel estimations. We have not stated nor argued in the text that this selection of distribution results in a reduction in terms of uncertainty. In addition, we stated (see section 2.4) that the selection of the Gumbel model is based on: 1) a likelihood-ratio test done by Dankers and Feyen (2008) where it was shown that the GEV model is only (statistically) justified in 15% of the grid cells, and 2) results where they found no persistence trend on estimations of the return levels based on GEV or Gumbel, therefore, being both equally valid. As highlighted in the text, we are aware of the implications of neglecting the uncertainty in the shape parameter (ξ) by assuming a Gumbel distribution. Results from Dankers and Feyen (2008) suggest that the Gumbel model is a good alternative for the domain studied in this work.

SC2. I dont understand Figure 10 and dont believe the r2 values if they are Nash-Sutcliffe, if they are simple correlation coefficients on the annual average discharges, the correlation still seems to be very high, what is EF.

A2. Series plotted in Figure 10 of the manuscript correspond to average discharge and average annual maximum discharge for each of the 554 stations in the period 1961-1990. We used model efficiency (EF) 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. These are **not** time series for the gauging stations. To avoid any confusion we are including the results as a complement to this response.

SC3. Figure 11 is of limited value in its present form, as the fit between observed and simulated values is not displayed in such a way that it is visible. I dont see the reason for plotting the biased values here, it is clear from the previous section that these will display dramatic differences compared to the observed values. In some instances the fit between the simulated and observed discharges (n.b. they are either levels or discharges, not both) is very poor (Guadiana, Danube, Themse, Nemunas, Daugava, Garonne, Rhone, Ebro, Kenijoki, Duero). A considerable number of the streams modelled in this study are in fact man controlled and thus the extreme discharges cannot be replicated without implementing the control strategy for the basin. This would obviously go beyond the scope of the study, however it is not mentioned and the way the diagrams are laid out could suggest that the poor fit in terms of extremes is deliberately hidden by the authors. I dont see the quality of fit as a major issue if it is acknowledged and relevant suggestions are made. Another point I find strange is the number of observed values in figure 10 and the return period of the largest value. If these are observed values and the data series are more extensive than the modelling period, then only values from within the modelling period should be used. For most large European basins the assumption of stationarity is definitely not valid over a longer time scale. Also plotting values up to return periods of 1 in 1000 are questionable based on data covering 30 years.

A3. We partially agree with this comment. For clarity we explain what is plotted in Figure 11 of the manuscript (reproduced here as an enhanced version, see Figure 1). 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. Return periods for the black crosses reach a maximum of ca. 70 yr given that we are working with 30 annual maxima. This representation of return levels using empirical plotting positions for observed data is employed in standard software packages dealing with extreme value analysis (e.g. extRemes and ismev), thus, we believe it fits our purpose.

These results come directly from the fitting of the Gumbel distribution to the annual maxima discharges obtained from hydrological simulations driven by uncorrected and bias corrected forcing climate data. The reason for this figure is to show the drastic differences at certain stations for the Gumbel fittings when LISFLOOD is driven by different climate data and to highlight (to a lesser extent) the uncertainty associated with the fitting of the Gumbel distribution (95% CI). Of course, this is limited to a reduced number of points, therefore, this is extended to the whole domain studied in this work in Figure 12.

As highlighted in the manuscript, we acknowledge that limitations in our approach may arise from "river regulation" (among others), and as stated by the referee, a further analysis on this is beyond the scope of the article. At the same time, Figure 11 has been designed in such a way to highlight the differences between both series of fitted distributions and not to deliberately hide the poor fitting of the Gumbel distribution as suggested by the referee. We simply do not support nor adhere to such practices.

As explained before, return levels for the observed discharges were obtained from the empirical plotting positions from the 30 annual maxima values (1961-1990) for each of the 20 gauging stations. We are not using additional time series.

In the revised manuscript we have clearly stated the limitations of the stationarity assumption for the error model.

SC4. I am not entirely sure what Figure 12 adds to the paper, surely the reader does understand by now, (if not already before reading this paper), that the use of uncorrected climate variations will result in considerable errors on both over and under estimation of average and extreme runoff values. Instead additional detail on the changes of extremes would be suitable such as looking at different return period events, e.g. 25 year versus 100 year.

A4. Figure 12 is the spatial representation of what is shown in Figure 11, which is only limited to 20 stations (river cells). This figure shows the ratio for Q_{100} between hydrological simulations driven by uncorrected and bias corrected climate data. It highlights that even if there is a strong overestimation of precipitation for the uncorrected climate data, this is not directly translated into an increase of Q_{100} compared to the simulations driven by bias corrected climate data. As explained in the manuscript, there exist complex interactions between snow cover, temperature and, possibly, actual evapotranspiration. We believe this figure adds valuable information to the reader. In addition, thanks to this figure (among others), we have discovered seasonal trends and relationships among different variables (e.g. 5-days precipitation, temperature, snow cover and number of frost days) that are interacting in the calculation of future Q_{100} estimations. These relationships have been observed for an ensemble of 12 climate models and this will be further explored in a forthcoming publication.

Technical corrections

TC1. Page 3894 top, should this be wet days per month, per season or per year?

A1. Corrected (see revised manuscript).

TC2. Figures 2, 4 5 & 6 are far too small to be readable and appreciated by the reader.

A2. It is challenging to fit in a full page a seasonal analysis for both uncorrected and bias corrected climate data for the whole European continent.

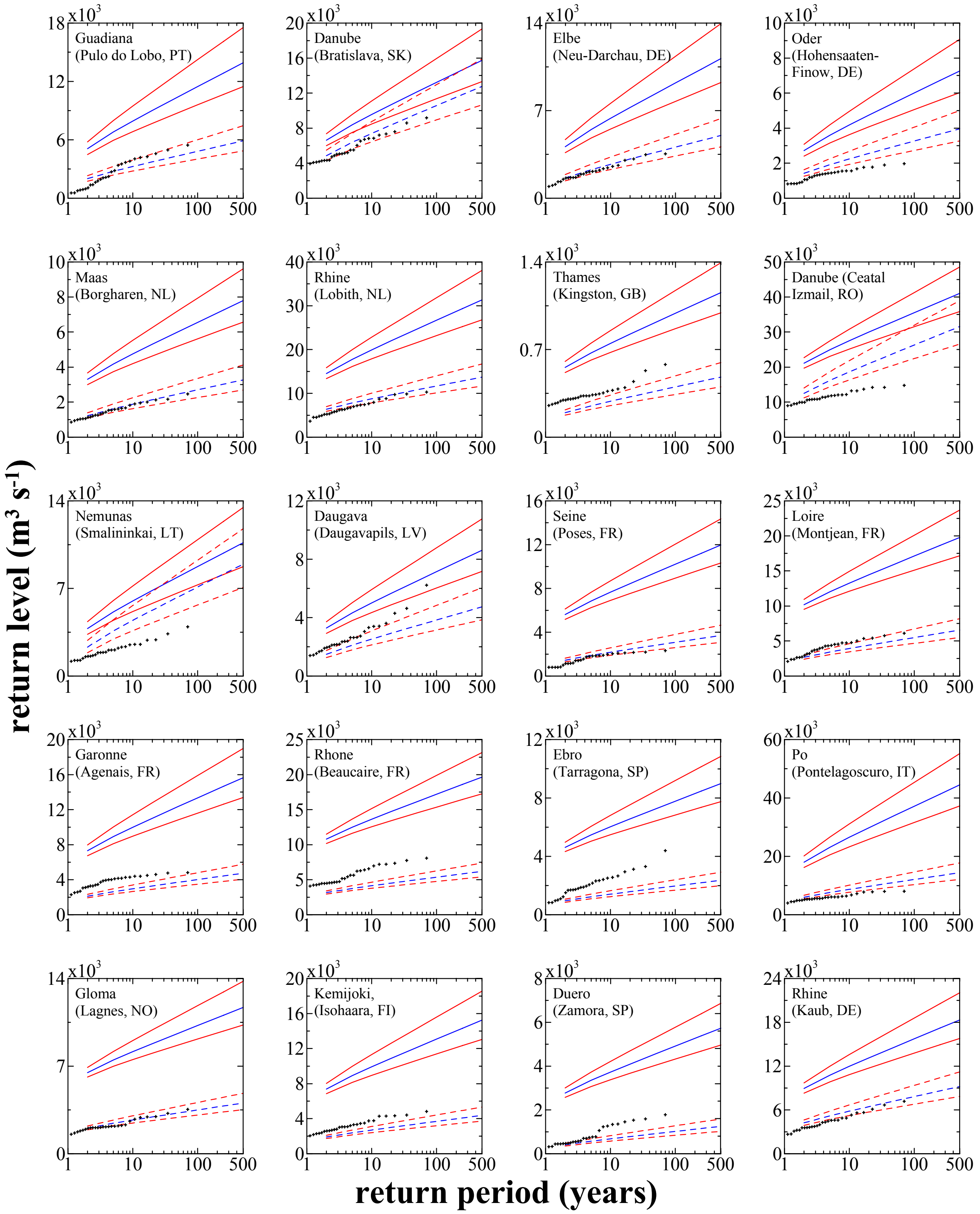

Figure 1: Return level plots of simulated discharge levels in the 20 selected gauging stations showed in Fig. 1, based on a Gumbel distribution fit to the annual maxima for the control period 1961–1990. Full lines represent hydrological simulations driven by uncorrected forcing data whereas dashed lines represent the bias corrected counterpart. Also included in the plates are the 95% confidence intervals (red dashed- and full-lines) derived using the profile-likelihood method. Black crosses represent return levels obtained from empirical plotting positions of observed annual maximum discharges at the selected stations.

Bibliography

Dankers, R. and Feyen, L.: Climate change impact on flood hazard in Europe: An assessment based on high–resolution climate simulations, Journal of Geophysical Research, 113, D19 105, doi: 10.1029/2007JD009719, 2008.

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.