

Actual edits made to the manuscript are written in red at the end of answers.

Interactive comment on “Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models” by I. Giuntoli et al.

I. Giuntoli et al.

ixg282@bham.ac.uk

Received and published: 20 March 2015

Anonymous Referee 1 Received and published: 21 January 2015

Comment-1 The manuscript covers the interesting topic of hydrological extremes with respect to changes in the 21st century. The impact and uncertainty related to these projections is discussed in the manuscript. Although the topic is of great interest and the writing is good, some issues need to be addressed before the manuscript should be accepted for publication. My biggest concern is the novelty of this work. How is this work related to other studies and related to other papers that also use ISI-MIP data

[Interactive
Comment](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

[Discussion Paper](#)



and GCM-GHM combinations (papers are cited in this manuscript) and how does the presented work differ from these other studies.

Answer We thank the reviewer for the helpful suggestions.

As pointed out by Vetter et al., (2015), numerous studies on climate change impacts have provided useful knowledge, however a synthesis of climate impacts for different regions and a consistent estimation of uncertainties is still missing. Our study goes in this direction: we conduct for the first time a systematic analysis, which enables an interpretation and comparison of results across space and hydrological regime.

We realize we have not sufficiently emphasized aspects regarding the novelty of the paper, which can be summarized in the following points:

- i) we investigate how projections in hydrological extremes are linked to the underlying climate and hydrological processes. For the first time this is done:
 - through a consistent joint estimation of extremes at once
 - at the annual and seasonal scales
 - at the global and regional scales using the Koppen-Geiger analytical framework.

We thus describe how high and low flows and inherent uncertainty vary at the seasonal and spatial scale, identifying areas where we have more confidence in the climate or in the hydrology (i.e. uncertainty is owed to GCMs or GHMs).

- ii) we assess the uncertainty using a formal statistical approach (2-way ANOVA model - having tested the model assumptions described in Appendix C) partitioning the variance using GCMs and GHMs as factors. Other papers that make use of the ISI-MIP data set (Schewe et al. (2013), and Dankers et al. (2013)) partition GCM/GIM uncertainty using ratios between the variances, while our ANOVA

approach adds the contribution of the error (or residual) to the partition of the variance. The uncertainty assessment by Prudhomme et al. (2014) omits the partition into GCM/GIM, providing a Signal-to-Noise (standard-deviation/mean and inter-quartile-range/mean) that does not allow for a gridcell by gridcell estimate of the major contribution to uncertainty.

- iii) We make a consistent joint estimation of the contributions to uncertainty for both extremes at once, having identified both extremes with the same method (5-d fixed-window variable threshold method). It should be noted that for high flows Dankers et al. (2013) have looked at a different metric (30 yr flood return levels of runoff of 5-d average peak flows), and for low flows we used an improved version of the Variable Threshold Method used by Prudhomme et al. (2014).

We will amend the manuscript (Introduction and Discussion sections) to reinforce these novelty aspects and to be more exhaustive on the relations and differences with other ISI-MIP based papers. [Edits listed in answer to Comment-2](#)

Major remarks

Comment-2 Page 4 Line 24- Page 5 Line 9: why is this analysis so different from Prudhomme et al (2014)? I see the point of the slightly different threshold, but Prudhomme et al also performed an analysis into the uncertainty of the GCM-GHM combinations for the low flows. Moreover, the results are nowhere compared with Prudhomme et al (2014). Are they really so different with this new (short moving window) threshold approach. Please convince me and the reader of the manuscript that this is really new work related to the low flow analysis. Furthermore it is stated on Page 11 Line 6-8 that other studies did study the high flow as well, so I have some problems with the novelty of this work. I see some novel aspect, but they are not clearly pointed out in the manuscript. The author point out 2 studies into future high flow and 2 studies into future low flow, of which some use the ISI-MIP data and

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

are all GCM-GHM combinations. Please convince the reader why this paper is novel compared to the other work, other than that it contains both hydrological extremes in one paper. Moreover, how the finding here differ from the other studies, in particular Prudhomme et al and Dankers et al.

Answer With respect to low flows, there are three major differences in the method and its implementation. Firstly, the index on which mean changes are computed follows the same concept of variable threshold method employed in the work by Prudhomme et al. (2014), but it has been revisited in order to overcome a limitation of the 30-d moving window for which gridcells were assigned lower threshold values than the theoretical value* (10% for percentile Q10), so a tendency to capture fewer occurrences, i.e. more extreme events (this effect is perhaps attributable to a slow emptying of reservoirs during the recession phase for some GHMs). The use of a 5-d fixed window eliminates this lower-than-theoretical-values effect – although the ensemble results for mean changes do not greatly differ (Prudhomme et al., 2014, Fig 1 with Fig 1-a of our manuscript). Secondly, in Prudhomme et al. (2014) the uncertainty analysis had been carried out grouping mean changes per type of model and computing the Signal-to-noise. This analysis only allows to infer which group (GCM/GHM) brings about highest agreement in the ensemble results; the partition of uncertainty we carry out in this paper moves beyond a signal-to-noise ratio: the quantification of each source (GCM/GHM) to total uncertainty via the sum-of-squares (ANOVA) allows to describe the spatial variability of the contributions gridcell per gridcell. Further, expressing the sum-of-squares of each source via the Koppen-Geiger regions allows for an improved understanding of how the climate and hydrological processes drive uncertainty for both runoff ends. Thirdly, the GHM JULES was excluded from the analysis because its grid-resolution differed from the one of the other GHMs, and also its runs induced considerably larger gridcell masking areas (see response to Comment 4). With respect to high flows, Dankers et al. (2013) have focused their analysis on annual extreme monthly flood peak (30-yr return level) as their aim was to describe flood

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



hazard changes.

The flood hazard patterns (Dankers et al. (2013) Fig. 1) show increases in northern North America, Eastern Asia, northern Latin America, Central/Horn of Africa, North-western China. The comparison of these patterns with our study (i.e. with the changes in the occurrence of high flow days – Fig 1-ab of our manuscript), reveals some similarities: mostly northern North America, and Northern Asia. However, in some regions, e.g. North-eastern Europe, we have opposite patterns. The uncertainty results, that Dankers et al. (2013) expressed with GCM/GHM variance (cf. Fig. 1 lower panels, Dankers et al. (2013)), is in agreement with our findings for the southern hemisphere, mainly driven by GCM uncertainty, whereas there is less agreement for the northern hemisphere (in North America, Central Canada is GCM driven uncertainty, whereas it is GHM driven in our results). It could be argued that the differences in the phenomena analyzed and the methods make the comparison between Dankers et al. (2013) and our work questionable. Thus the findings comparison must consider the differences between the two studies: i) the scope: change in flood hazard vs. change in frequency of high flow days; ii) the method: different metric and uncertainty partition framework (the variance is partitioned into two factors – GCM and GHM – while our framework adds the error as contributor to the variance); iii) the multi-model ensemble: it comprises the same GCMs (5), but uses 3 additional GHMs – JULES, LPJmL, MATSIRO that we did not use as they are not suitable for low flows (LPJmL and MATSIRO showed large areas with very low values hampering the extraction of the low flows index (i.e. hefty cell-masking), to a lesser extent JULES, though the latter was also excluded for grid-resolution issues as mentioned above).

We will better organize these aspects in the manuscript, clearly writing the differences with Prudhomme et al. (2014), and also with the methods and findings from Dankers et al. (2013) in the Introduction and Discussions Sections.

Introduction: P4 L.20 - P6 L.3 ; **Disc. Conclusions:** P12 L.1-5, L.13-20.

Comment-3 Page 4 Line 25 Why are two different threshold selected for the high and low flows. Why take the top 10% (extreme) for low flows and 5% (exceptional) for

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



the high flows. It is well known that GHMs and GCMs have difficulties reproducing the extremes, which is even truer for the GCM-GHM ensemble mean. When two different thresholds are applied the two cannot be directly compared and the same is true for the patterns and magnitude of the changes. Most likely the GCM-GHM combinations will have a different performance/skill in reproducing and projecting the 10th or 5th quantile. Additionally I found an abstract of the authors at AGU that uses only the 10th and 90th percentile, why did the authors change this later to 5th? Moreover, what is the impact of the selected threshold level on the uncertainty or SN2 ratio. Would a different threshold level result in different SN2 ratios or are the results stable and not threshold specific?

Answer There are justifiable physical process based reasons for selecting different threshold levels for low and high flow extremes. The two phenomena (high and low flows) are quite different: low flows are generally characterized by a slower onset, and a longer duration; while high flows by a sudden onset, and a shorter duration. They are not necessarily symmetric with respect to the median flow. In order to capture the intrinsic differences between the two phenomena we tested Q10 and Q15, and Q90 and Q95. Finally – for low flows: seeking a sufficiently low quantile without compromising the analysis (quantiles lower than 10% become intractable for the large presence of zero pools in some time series for which larger areas of the globe would be screened out) we chose the same as Prudhomme et al. (2014); for high flows we chose the 95th percentile. We had written the abstract for AGU2013 before carrying out the analysis with different quantiles and making the final choice, we then presented results at the fall meeting with the Lowflows Q10 Highflows Q95 setup as in the paper. As the reviewer points out the skill varies changing the quantile, this is especially true for low flows, for which the lower the quantile, the lower the skill (e.g. Gudmundsson et al. (2011) found that the performance of a similar set of global models – WaterMIP decreased systematically from high Q95 to low Q5 runoff percentile over Europe).

Comment-4 Page 5 Line 5: JULES was left out of the analysis, how much does this impact your estimates of the uncertainty compared to Prudhomme et al?

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Appendix B: P17 L. 13-24

Answer JULES inclusion would have increased the uncertainty for: i) its peculiar setup (inclusion of varying CO₂ and vegetation dynamics processes, absent in the other models) which make it an outlier compared to the other GHMs, and ii) its runs have a coarser resolution, so in the regridding process (to match the other GHMs in the Ensemble) from 1.8*1.25 to 0.5*0.5 resolution may influence uncertainty (gridcell fractionation of the same information into smaller sub-gridcells). The first point is quantified in Table S2 of Prudhomme et al 2014, in which the uncertainty (expressed by S2N ratio) is provided on e.g. column 4 for both the entire ensemble ("All GCMs GIMs") and for the ensemble minus JULES ("6GIMs"): the S2N is higher (stronger agreement between the models, i.e. decreased uncertainty) when JULES is not included: S2N@6GIMs 3.24 > S2N@AllGIMs 1.67. To assess the second point - i.e. the degree of impact of coarse to finer resolution on uncertainty - one would need to carry out the analysis using the same model setup and forcings at both resolutions; however, JULES runs at 0.5 are not available. This aspect is currently being investigated by the scientific community (Bierkens et al., 2015) seeking to describe the effects of resolution on model uncertainty. In essence, the inclusion of Jules would have driven the uncertainty towards the GHM source. As an illustration, the uncertainty maps of Fig. 1, Fig 3, Fig 4 would have been less yellow (GCM-driven uncertainty) and more green (GHM-driven uncertainty). During our exploratory analysis we have included Jules in the ANOVA, results of the uncertainty partition with and without JULES are shown in Fig. 1 and Fig. 2 of this document respectively – though test assumptions on the adequacy (i.e. residuals normality and constancy of variance) of the ANOVA model including JULES were less met. It should be noted that, for low flows, JULES inclusion (as seen in Fig. 1 and Fig. 2) has an effect on the masking which is homogenized across all GHMs and for which large areas of the world are screened out as its runoff time series presented large pools of zero values (similarly to LPJmL and MATSIRO).

Disc. Conclusions: P15 L.8-24

Comment-5 Page 6 Line 25-27: Gridcells with no seasonal changes are removed from the analysis. In which period (1970-2000 or 2066-2100) should these gridcells show no seasonal change? Moreover, with a changing climate gridcells on the edge of

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



the Sahara or Greenland can become seasonal. This in itself is an interesting finding. To follow up on this an analysis is done for different climate regions. However, these change over time as well and some are more likely to change than others. This was not taken into account and will impact the results.

Answer Gridcells removal is based on the control period (1971-2005). This will be specified in the Data and Methods section and in the Appendix B where we describe the gridcell masking procedure.

With regards to the gridcells potentially becoming seasonal with a changing climate, we feel the frontier no-change/change is difficult to follow at this spatial scale of analysis (i.e. global). As an example, in the literature Alessandri et al. (2014) have looked at the expansion and retreat of specific climate boundaries (Mediterranean climate in Europe and western USA) using CMIP5 climate projections. This aspect is very interesting but we think it lies beyond the scope of our study and it would require focusing on limited areas alone. [Data and Methods: P7 L.19-20; Appendix B: P18 L.26-29, P19 L.1-2](#)

Specific remarks

Comment Page 3 Line 6 mentions multi-model ensembles. However, the manuscript continues with single model studies (e.g. Sheffield and Wood 2008). The reader expects a summary of multi-model simulations; maybe move this line to later in the manuscript when multi-model studies are discussed.

Answer We will address this in the manuscript by moving (in Page 3) the block of Lines 6-to-10 to Line 20, just before the introduction of the model inter-comparison projects.

Comment Page 4 Line 19-20 I believe the sentence should be ... all GCMs for each GCM averaged over all of the GCM or am I mistaken?

Answer As described by the authors (see Schewe et al. (2013), caption of Fig. 1, and Dankers et al. (2013), caption of Fig. 1), the GCM variance of the change is computed across all GCMs for each individual GHM, and then averaged over all of the GHMs.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Comment Page 7 Line 11-13: This $SN2 > 1$ is rather arbitrary. Could there be a more statistically proper way to define this $SN2$ threshold?

Answer $SN2 > 1$ was chosen to facilitate the visualization of the model agreement on the maps. In our case, we wanted to remain consistent with Prudhomme et al. (2014) to facilitate comparisons.

Comment In general: Which dataset was used to computer the Koppen-Geiger climate classifications? I can assume they are different between the GCMs and different datasets. Please specify in the manuscript.

Answer The dataset we used for the Koppen-Geiger climate regions is based on present-day following the global data classification from Kottek et al. (2006) (a link is provided at the end of Table 1 in the manuscript - http://koeppen-geiger.vu-wien.ac.at/pdf/kottek_et_al_2006_A4.pdf). Therefore, there is no difference between models: all gridded model outputs have been assigned a Koppen-Geiger climate region from the Kottek et al. (2006) dataset. We will clarify this in the manuscript, in particular in the Data and Methods section (Page 8, Line 7). We are aware that region boundaries can change with time (as investigated by e.g. Wanders et al. (2014)); however, in our study we were seeking changes in projections with a present baseline as reference. **Data and Methods: P8 L.27-28**

References

Alessandri, A., De Felice, M., Zeng, N., Mariotti, A., Pan, Y., Cherchi, A., Lee, J.-Y., Wang, B., Ha, K.-J., Ruti, P., and Artale, V.: Robust assessment of the expansion and retreat of Mediterranean climate in the 21st century, Sci. Rep., 4, <http://dx.doi.org/10.1038/srep07211>, 2014.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A.,

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

D. V. Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, *Hydrological Processes*, 29, 310–320, 10.1002/hyp.10391, <http://dx.doi.org/10.1002/hyp.10391>, 2015.

Dankers, R., Arnell, N. W., Clark, D. B., Falloon, P. D., Fekete, B. M., Gosling, S. N., Heinke, J., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble, *P. Natl. Acad. Sci. USA*, 111, 1–5, 10.1073/pnas.1302078110, 2013.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, *Journal of Hydrometeorology*, 13, 604–620, <http://dx.doi.org/10.1175/JHM-D-11-083.1>, 2011.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, *Meteorol. Z.*, 15, 259–263, 2006.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment., *P. Natl. Acad. Sci. USA*, 111, 3262–3267, 10.1073/pnas.1222473110, 2014.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, *P. Natl. Acad. Sci. USA*, 111, 3245–3250, 10.1073/pnas.1222460110, 2013.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., and Hattermann, F.: Multi-model climate impact assessment and intercomparison for three large-scale river basins on

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



three continents, Earth System Dynamics Discussions, 5, 849–900, 10.5194/esdd-5-849-2014, 2014.

Wanders, N., Wada, Y., and Van Lanen, H. A. J.: Global hydrological droughts in the 21st century under a changing hydrological regime, Earth Syst. Dynam. Discuss., 5, 649–681, doi:10.5194/esdd-5-649-2014, 2014.

Interactive comment on Earth Syst. Dynam. Discuss., 6, 1, 2015.

ESDD

6, C1–C13, 2015

[Interactive
Comment](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

[Discussion Paper](#)



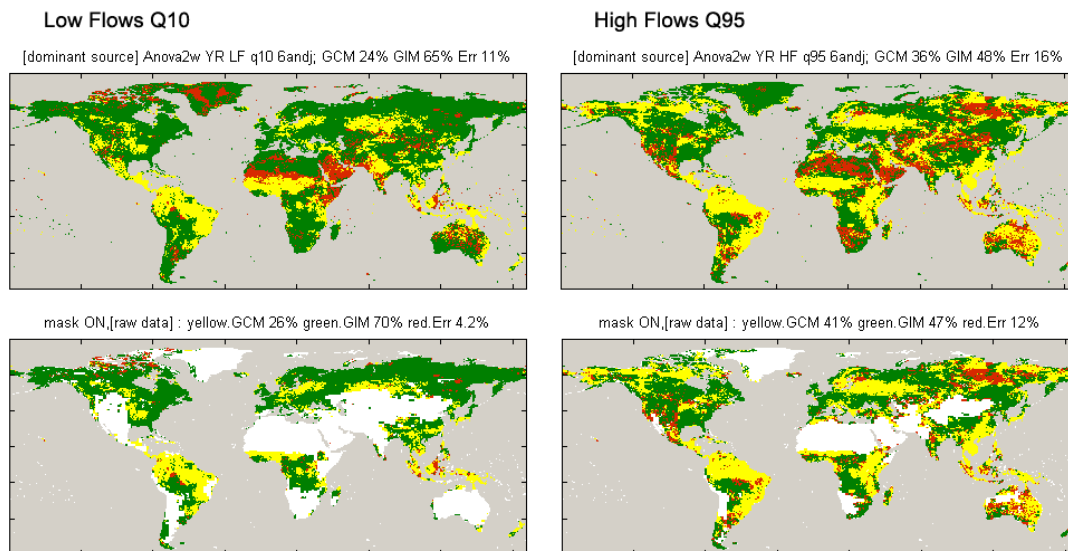


Fig. 1. Uncertainty partition - including JULES GHM

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

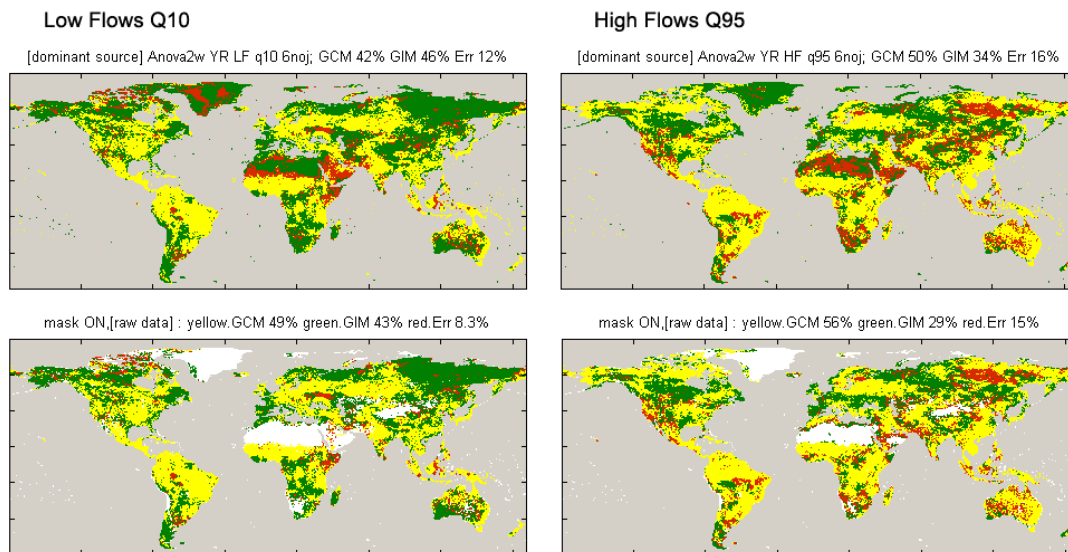


Fig. 2. Uncertainty partition - excluding JULES GHM

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

Interactive
Comment

Interactive comment on “Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models” by I. Giuntoli et al.

I. Giuntoli et al.

ixg282@bham.ac.uk

Received and published: 14 March 2015

Anonymous Referee 2 Received and published: 2 March 2015

This paper is a concise analysis of the potential consequences of projected climatic conditions on low-flow and high-flow frequencies globally. On the whole, the paper is very clearly written, and I only have a few concerns.

Answer We thank the reviewer for the positive review and helpful suggestions.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Comment-1 As Section 4 of the paper correctly notes, there are several sources of error not investigated in this study, such as bias correction, CO₂ and vegetation dynamics, emission scenarios, and internal variability. In view of these sources of uncertainty, the abstract (line 21) should not state that using multiple GCMs and GHMs is sufficient to envelop the overall uncertainty range.

Answer We agree, the use of multiple models alone does not allow embracing all of the uncertainty. We will edit the text accordingly. **Abstract: P2 L.21-22**

Comment-2 Page 4, lines 8-15 also bear on the issue raised in my first comment. Prudhomme. et al. (2014) and Davie et al. (2013) are reported to find that biome models which include effects of varying CO₂ produce more runoff than purely hydrologic models. This implies that using GHMs without a varying CO₂ effect not only will not envelop the overall uncertainty range, they may also bias the results toward less runoff. So some runoff possibilities may be missed by a set of hydrological GHMs, and at the same time some erroneously low runoff solutions might be improperly included within the uncertainty range.

Answer We agree, this is a critical point. The aim of including a rich set of GHMs in order to capture a representative uncertainty range had to come into terms with the applicability of the method. During our analysis the two biome models available through ISI-MIP have shown the presence of large pools of zero values in their time series that caused heavy land masking, thereby hampering the global comparison assessment. As shown in Fig.1 index extraction was vetoed over large parts of the globe leaving only 61% for JULES and 20% for LPJmL of land gridcells after masking (with the veto rules described in Appendix B). Note that the masking applied to the manuscript's ensemble is formed by superimposing masking from each of the GHM-GCM combinations.

The unavailability of historic simulations with the same GHMs to evaluate the GHMs' ability to represent the different aspects of runoff regime (e.g. high, medium and low flows) lead us to adopt the pragmatic approach of using the largest possible ensemble

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

of models common to both high and low flows. Another option could have been to use one ensemble for low flows and one for high flows (with JULES, though posing limitations with resolution and masking), but we would lose consistency in the global comparison of projections of changes and ANOVA across flows.

We are aware of the potential underestimation of the uncertainty or bias towards non-biome GHMs, and this is not confined to CO₂ processes in plants but also include spatial resolution (as the reviewer mentioned below), internal variability in GCMs, GCMs and GHMs parameterization, and lack of detailed processes such as e.g. surface and groundwater exchanges or glacier in GHMs. However, our results demonstrated that, even excluding biome models and other model structure differences in the ISI-MIP ensemble, large uncertainty in the signal of changes in high and low flows is attributable to GHMs and not only on GCMs. We will amend the abstract and extend the discussion to highlight this important point. We will also attach to Appendix B a table of gridcell availability per GHM-GCM combination for high and low flows.

Introduction: P4 L.11; **Disc. Conclusions:** P15 L.6-29; **Annex Table A2:** P27

Comment-3 A related issue not mentioned, but which should be mentioned, is that the mere inclusion of multiple models is insufficient to fully scope the uncertainty associated with models. Structural model errors are not all random; some structural errors, such as insufficient resolution, are common across all models. Using multiple models does not help with this aspect of model uncertainty.

Answer We agree, we will refer to this aspect of the difficulty to fully scope the models uncertainty in our revised manuscript (as in our reply to Comment-1).

Disc. Conclusions: P15 L.24-27.

Comment-4 Page 5, bottom should include mention of which (if any) GHMs include varying CO₂ effects. At present, the reader is left to wonder until near the end of the manuscript.

Answer Yes, as suggested, we will state in the Data and methods Section that none of the GHMs used in our ensemble include varying CO₂. **Data and Methods:** P6 L.25-26

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Comment-5 Page 6, lines 21-25 are written incorrectly. As written, for HFD, a cell exceeding the Q95 value may be assigned a value of either 0 or 1, while the assignment for a cell that doesn't exceed the Q95 value is undefined. The same applies to LFD. Lastly, the reference to the Appendix should be to Appendix B.

Answer Yes, actually the value of 1 is assigned only if the cell exceeds the Q95. We will rephrase: "For HFD, a value of 1 (high flow) is assigned to each cell if the cell's runoff exceeds the Q95 value, otherwise a value of 0 (no high flow) is assigned"; and similarly for LFD. [Data and Methods: P7 L.14-18](#)

Comment-6 Page 8, line 24 to Page 9, line 4: Seasonal differences are to be expected, but there seems to be no obvious reason why the NH in boreal winter should behave like the SH in austral summer, or why the NH in boreal summer should behave like the SH in austral winter. Yet, by discussing globe-wide variations in terms of DJF and JJA, this is what you are implying. To better frame the discussion, and to see whether this surprising possibility is borne out by the data, create separate Fig. 2 charts for each hemisphere.

Answer Approximately 80% of landcells are located in the NH, while 20% in the SH (not considering masked areas). Therefore, global results are dominated by boreal seasonality (high flow changes dominant in DJF, and low flow changes dominant in JJA). Splitting the manuscript's figure into three latitude ranges (Fig.2): North ($\text{Lat} > 23.5$); Tropics ($23.5 < \text{Lat} < -23.5$); and South ($\text{Lat} < -23.5$); the North still accounts for the majority of land (65%) with results that are broadly similar to the global ones. With three latitude bands, we can notice PDF shifts of e.g. DJF towards decreased changes for high flows and increased changes for low flows.

[Figure 2: P29; Results: P9 L.23-27](#)

Interactive comment on Earth Syst. Dynam. Discuss., 6, 1, 2015.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

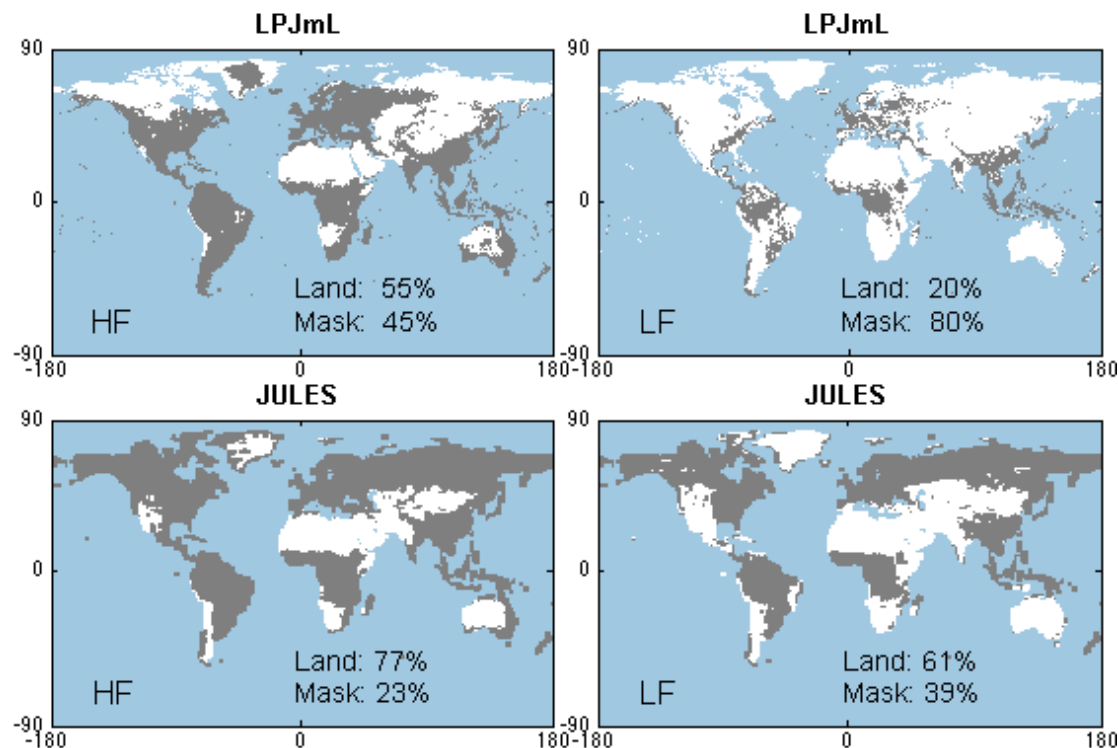
[Interactive
Comment](#)

Fig. 1. Land masking upon HFI and LFI index extraction for the biome-model GHMs (not included in the ensemble): LPJmL and JULES (top and bottom panels respectively).

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

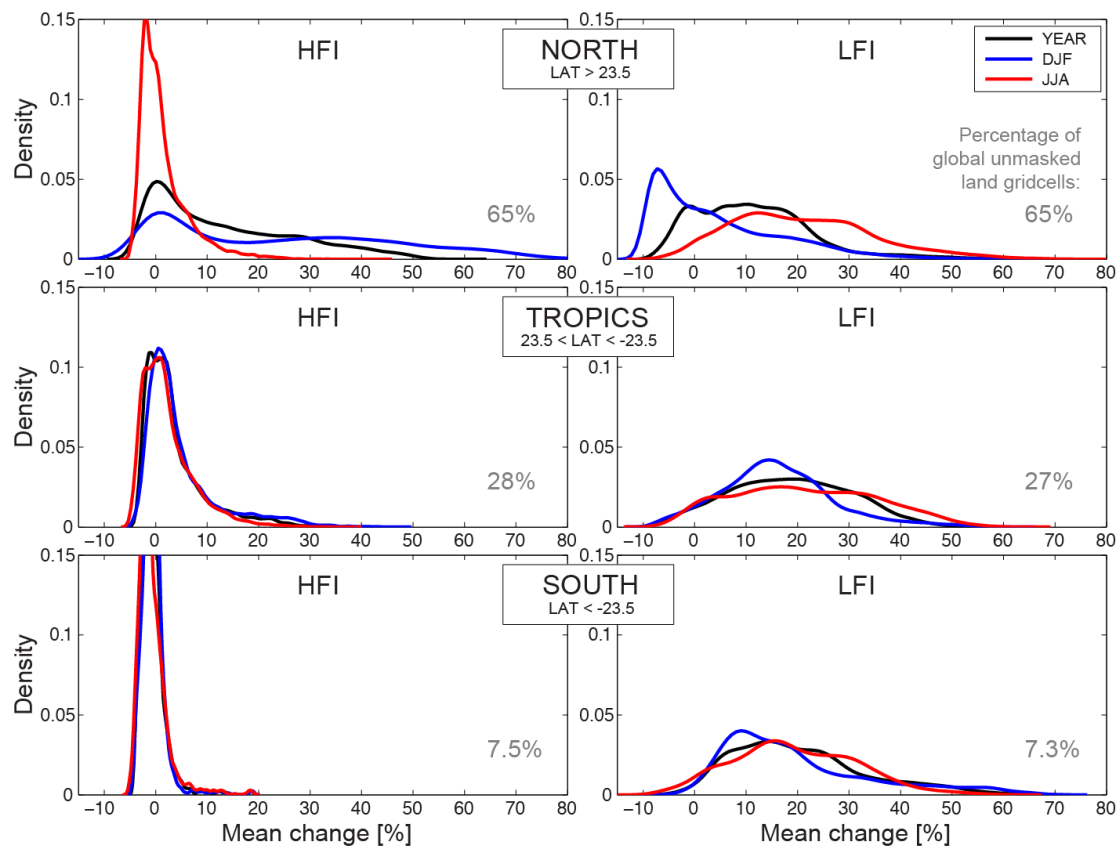
[Interactive
Comment](#)


Fig. 2. PDFs of mean changes in high (HFI) and low (LFI) flows, annually and per season (DJF and JJA) referring to North, Tropics and South latitude bands.

[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)
[Discussion Paper](#)


Interactive comment on “Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models” by I. Giuntoli et al.

I. Giuntoli et al.

ixg282@bham.ac.uk

Received and published: 14 March 2015

Referee 3 Antonio Speranza Received and published: 2 March 2015

Comment-1 The addressed topic is interesting and quite a bit of performed work is reported in the paper; in particular, I appreciated the effort in assessing the relative contribution of GCMs and GHMs to “uncertainty”. I think, however, the reader could be helped in forming his own judgement if some more information was provided here and there; for example, by showing specific examples (giving an idea of the nature of analysed fields) of runoff fields for different models.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Answer We thank the reviewer for the positive and helpful suggestions. As suggested, we could provide in Annex a specific example of mean annual runoff from the control period for five representative gridcells (Fig.1) i.e. one per main Koppen Geiger region (A-Tropical, B-Arid, C-Temperate, D-Cold, E-Polar) for the different global model combinations i.e. GCMs-GHMs. [Annex Figure B2: P35](#)

Comment-2 At some points I had problems in following the proposed reasoning:

- The proposed analysis suggests major impact of GHMs on polar and arid areas, where the runoff component is minor. Doesn't this go without saying?
- In Fig. 1 to the observed increase in the percentage of HFI days in high latitudes in the northern hemisphere (Fig. 1a) values of signal-to-noise ratio less than 1 correspond; only in some regions - Alaska, north-west America, north China - $S2N > 1$ (model agreement): what motivates, then, the conclusion "...exhibit a number of robust large-scale features. Increases in high flow days were found at northern latitudes, with a strong signal over..."? Similar considerations hold for seasonal maps.

Answer 1st point: True, however, in that instance we thought it was worth specifying it to contextualize how GHMs, which are ultimately responsible for simulating runoff over land (unlike GCMs), are the ones associated with large uncertainties in those areas where runoff is scarce and/or ephemeral - and therefore its simulation is difficult to achieve.

2nd point: We feel the $S2N \geq 1$ regions over northern latitudes are quite large and well clustered together, although we realize there are areas where regional agreement is weak ($S2N < 1$) as in e.g. northern Russia or western North America.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

Comment-3 I share with Referee 1 doubts concerning the “novelty” of proposed work, but I leave to him evaluation of answers proposed by the Authors.

All in all, I believe the paper is in line with similar literature and it can be published with the suggested amendments.

But, just for the sake of scientific discussion, let me raise a general question concerning the specific use of observations over threshold proposed in this paper.

The authors mention at pag. 12, lines 15–29 that “The identification of high and low flows over long time series, and particularly over climate projections, is non trivial. As an illustration, van Huijgevoort et al. (2014) in their multi-model ensemble study on droughts, report that applying the threshold level method to the future period using a threshold derived from the control period can lead to spurious pooling of drought events. They suggest that future changes could be counted for by linking the drought threshold to adaptation scenarios like Vidal et al. (2012) did over France. Wanders et al. (2014) used a transient threshold level method for a moving reference period, in order to reflect the changes in hydrological regime over time, finding that the non-transient threshold method projected larger shares of areas in drought (except in snow dominated regions).”.

Now, use of very low (because of trend) thresholds for the selection of “extremes” generates problems typical of “fat distributions”: the sets of events over threshold are rich in number, but include non extreme events! Use of different models further complicates the situation. As a consequence, even under conditions allowing application of Gnedenko (“three type”) theorem*, the distribution of the selected events is non- parametric and the reliability of the statistical inference is very poor (in particular for what concerns the “tail events”). But, at the same time, “central limit” conditions are not fulfilled and the distributions cannot be considered normal*. Geographical non-uniformity further complicates the picture.

As a consequence, the inferred statistical estimators are presumably characterized by

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

(very) weak reliability (in particular for what concerns “real extremes”) and have, at most, a qualitative meaning: I would not base any relevant decision on them!

*Note that in the specific case in question even a simply linear trend would introduce an additional dimension to the parametric space of the distribution.

*By the way, isn't normality a necessary requisite in ANOVA?

Answer The discussion you raise is very interesting. The problem of fat distributions upon selection of low flows extremes is well known when dealing with hydrological time series. A study on a global scale has to come into terms with a threshold selection (in our study Q10 was selected over 34 years on a 5-d basis) that will have to be universal and serve many hydrological regimes and climate regions at once i.e. geographical non-uniformity. A threshold selection resulting from this trade off, will describe low flows but will inevitably pool events that are not truly extreme. Nevertheless, we simply looked at if the threshold was crossed more or less often in the future with respect to the past, therefore the technical issues of poor reliability of statistical inference should not be as pressing as in the context of frequency analysis (i.e. fitting extreme value distributions).

With regards to the ANOVA we carried out in our study, we tested the residuals for normality, and, although not always, this condition was met by 78% and 88% of unmasked land gridcells for HF and LF respectively (Appendix C). [Appendix B: P17 L.13-16](#)

Interactive comment on Earth Syst. Dynam. Discuss., 6, 1, 2015.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



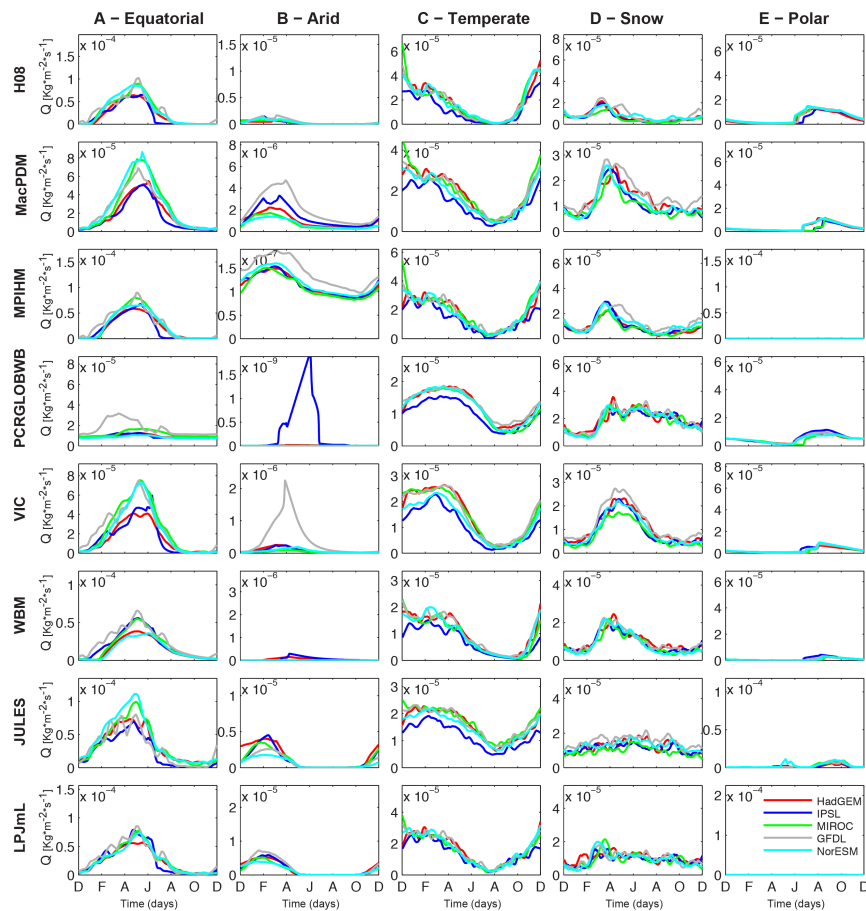


Fig. 1. Mean annual daily runoff (1972–2005) per GHM (in row) for selected gridcells (Lat, Lon): A-Tropical (-2.25, -53.25); B-Arid (-20, 25); C-Temperate (43.75, 11.25); D-Snow (41.65, -91.5); E-Polar (65, 165).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Interactive
Comment

Interactive comment on “Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models” by I. Giuntoli et al.

I. Giuntoli et al.

ixg282@bham.ac.uk

Received and published: 16 March 2015

Anonymous Referee 4 Received and published: 3 March 2015

General Comments The authors present a study that investigates the projected changes in hydrological extremes from 6 global hydrological models. The study is robust, both in terms of the statistical techniques used and the size of the ensemble (although the authors can easily extend the number further). The results of the study further our understanding of uncertainties global climate and hydrological models in simulating in future hydrological extremes.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Answer We thank the reviewer for the positive review and helpful suggestions.

Comment-1 My main concern is with the authors' use of multi-model ensemble mean. Although effort has been made to provide some information on the inter-model spread (e.g. variance), the manuscript can make a significant contribution to our understanding of the projections and modelling uncertainty of each GHM-GCM combination. In these respect, Figures 1, 3 and 4 can be provided as added illustration in the manuscript (or as appendix) by showing the changes in frequency of days under high and low flows from every GHM-GCM combination. In addition, given the focus of the study, I have strong recommendation on the use of winter half-year and summer half-year periods which is more representative of the relationship between rainfall and runoff, instead of climatologically defined seasons such as DJF and JJA.

Answer We understand the reviewer's concern with the use of the multi-model ensemble mean, but we must keep in mind that the main goal of this study is to provide a comprehensive synthesis of multi-model ensemble hydrological projections and variance within (i.e. uncertainty). Considering single GCM-GHM combinations would mean annexing 60 mean changes maps (30 for high and 30 for low flows) on maybe 5 or 6 A4 pages. It's probably too much space devoted to the behavior of each model combination given the core focus of this study.

We acknowledge the value of using winter half-year and summer half-year periods, however, having looked at the four seasons (DJF, MAM, JJA SON), our choice went for seasons for which changes are more marked (as shown in Fig.1). This time scale also facilitates direct comparisons with similar studies. **Annex Figure B3 - B4: P36-37**

Specific Comments The authors should also address the following comments for added clarity to improve the manuscript.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

Comment-2 3:13-15 : The authors cited several studies which assessed future changes in the global water cycle and argued that the lack of GHMs in these studies presents a limitation. The authors went on to say that GHMs provide more uncertainty. What is the authors' point on the relationship between the utility of more GHMs and uncertainty? Are the authors referring to the range of uncertainty (i.e. ensemble spread) that is produced given the use of large number of GHMs? This should be made clear. In addition, the authors should explain: the types uncertainties by GHMs and if there are merits in using GHMs (with input from GCMs) compared with using solely GCMs for assessment of future hydrological extremes.

Answer As the reviewer points out, when considering the inclusion of more GHMs in relation to uncertainty we refer to capturing a comprehensive range of uncertainty. Terrestrial water cycle processes in GCMs can be simulated at a coarser resolution and with less detailed processes than in GHMs. This is why a cascade of GCMs-GHMs model ensembles are currently used to undertake global climate impact studies in hydrology. As detailed in our reply to Ref.2, our text relating to the uncertainty captured within the ISI-MIP ensemble will be revised in the new manuscript.

Data and Methods: P8 L.6-7; Disc. Conclusion: P15 L.6-29

Comment-3 4:12-13 : The cited studies – Dankers et al. (2013), Schewe et al. (2013), Davie et al. (2013) and Prudhomme et al. (2014) – do not amplify how “GCMs and GHMs uncertainty contribute to the spread in projected changes in hydrological cycle”. Reference to these studies should describe the regions where there is consistency as well as uncertainty in the projected changes of hydrological extremes.

Answer This has partly been addressed in the Discussion and conclusion Section, but we will make sure we have a comprehensive reference to these studies as of where the results agree i.e. changes in consistency and uncertainty occur.

Disc. Conclusion: P12 L.1-5, L.13-20

Comment-4 5:9 : Is there any reason why the authors use the Koppen-Geiger climate classification over other classifications such as Giorgi-Francisco regions (2002) or

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

Mahlstein-Knutti's (2010) cluster analysis-based regions?

Answer The aim of using regional classifications was to sort land (gridcells) into homogeneous regions to help summarize and examine results regionally. Giorgi-Francisco (2002) and Mahlstein-Knutti (2010) regions could be an option too, but we opted for the Koppen-Geiger classification because it is based on hydro-climate homogeneity, and it is a well known and referenced classification, which we thought would provide the best regional frame of reference for our study.

The Mahlstein-Knutti (2010) regions are very similar to the Koppen-Geiger ones (which are also based on temperature and precipitation), but their partitioning, in addition to present, also incorporates projected changes in the simulated variables (using CMIP3 models). The adoption of the MK2010 classification, would require to update the region boundaries by undertaking a new regional clustering this time using the data set at hand (i.e. CMIP5 models).

Comment-5 5:19 : What variables from GCMs were used as inputs to GHMs?

Answer The GHMs vary in the use of all or some variables as input, as listed in Table 1 of the manuscript "Meteorological Forcings". They include: Surface air temperatures (Tavg, Tmin, Tmax); Precipitation (snow and rainfall separately); Surface radiation (short and longwave downwelling); Near-surface windspeed (total); Surface air pressure; Near-surface relative humidity; CO2 concentration. The last two variables have not been bias corrected. All of the variable have daily and monthly frequency except CO2 concentration (annual).

These variables, as reported in the ISI-MIP Protocol, consist of bias-corrected climate data from the GCMs participating in the CMIP5 and cover the time period from 1950 to 2099 (1950-1970 are usually used for spin-up). **Appendix A: P16 L.24-27 - P17 1-2**

Comment-6 5:22 : There should be a description on how well the climates of CMIP5 GCMs have been bias-corrected. For example, can the authors identify regions where

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



there is high confidence in the simulation of climate after bias correction? The same can be said for regions which still suffer from poor simulation of climate after bias correction. (cf. 6.26)

Answer The regional variability in the performance of bias-correction is an interesting aspect of climate change science but we don't think it is relevant to our analysis because all GHMs were driven by the same bias-corrected climate forcing. In addition, more than one variables have been bias corrected and fed to GHMs, so the confidence in the simulation after bias-correction will vary according to the variable considered. We provide reference to the published material regarding the bias-correction method and invite the interested reader to refer to this body of work (Hempel et al., 2013).

Comment-7 6:21-24 : The binary assignment – 0 or 1 – to no low/high flows and low/high flows is confusing. In the example of high flows, if the runoff value of the cell exceeds the Q95 value, then the cell is assigned either 0 or 1. There should be separate criteria to distinguish a 0 flag from a 1 flag.

Answer This issue was also raised by Ref.2. As replied to Ref.2, we will rephrase the binary assignment as follows: “For HFD, a value of 1 (high flow) is assigned to each cell if the cell's runoff exceeds the Q95 value, otherwise a value of 0 (no high flow) is assigned”; and similarly for LFD. [Data and Methods: P7 L.14-18](#)

Comment-8 6:26 : On the aspect of “screened-out” areas – While the screened-out gridcells are located in arid or frozen regions, could they also be regions where there is considerable modelling uncertainty GHMs and/or GCMs? (cf. 5:22)

Answer Areas have been screened out on the basis of veto rules (described in Appendix B) applied to the HF and LF threshold curves (e.g. gridcells with null variance). Can these areas coincide with regions with considerable uncertainty i.e. considerable variance? They can, because the screening indicates large pools of zero-runoff but one GHM is sufficient to have that to impose the screening to the whole

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



ensemble, while the remainder of the GHMs could show considerable variance. In general however, the screening pattern is similar across all model combinations.

Comment-9 8:13 : Referring to Figure 2, the authors state that “the mean change vary spatially in magnitude”. However, there is no information in Figure 2 to highlight spatial differences. In addition, it is not clear how the mean changes are derived. Are these spatially averaged changes? Can the authors clarify these two points?

Answer This figure was crafted to provide a visual on the magnitude of mean changes that are mapped in e.g. Fig.1a of the manuscript. As written on the reply to Ref.2 we will update Figure 2. We provide in this document an updated version (Fig.1), which includes the seasons MAM and SON, in support to our reply to comment-1 on the choice of sub-annual windows of analysis. [Figure 2: P29; Results: P9 L.23-27](#)

Comment-10 8:14-19 : The description in the lines 14-19 should be referenced to Figure 1.

Answer We will amend as suggested.

References

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, *Earth Syst. Dynam.*, 4, 219–236, doi:10.5194/esd-4-219-2013, 2013.

Interactive comment on *Earth Syst. Dynam. Discuss.*, 6, 1, 2015.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



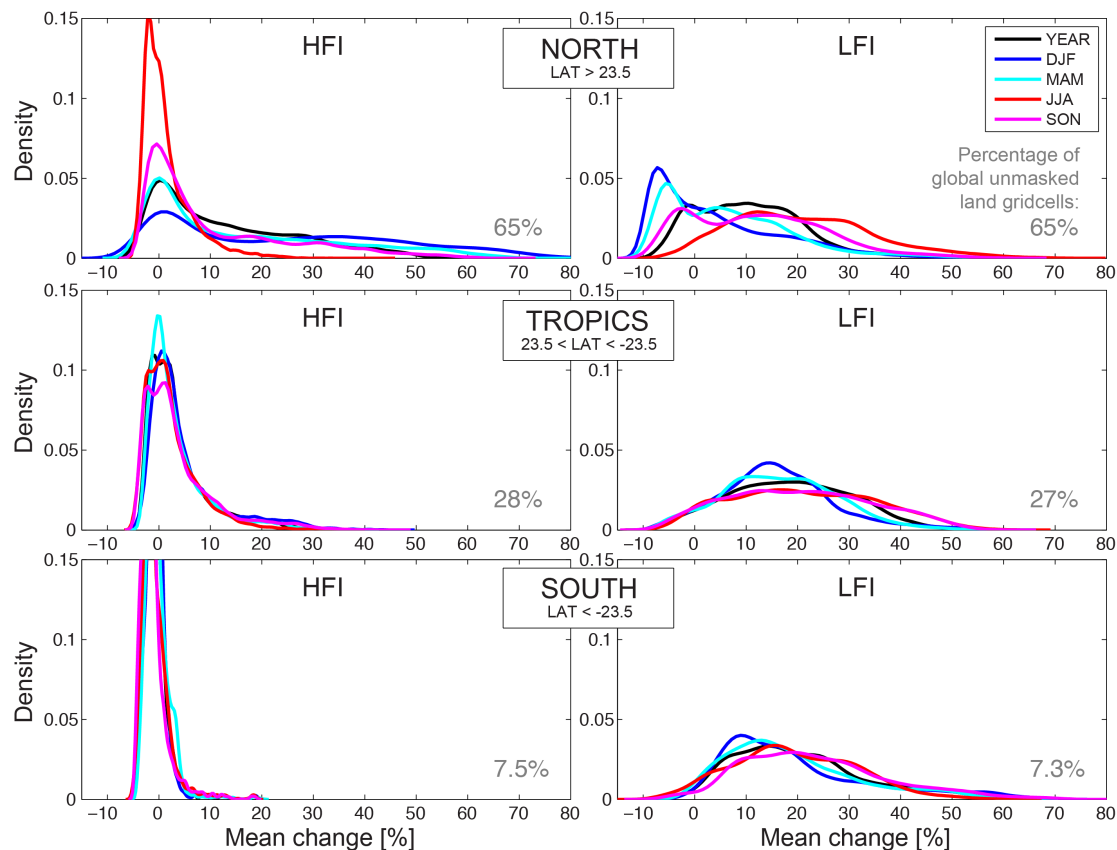
[Interactive
Comment](#)


Fig. 1. PDFs of mean changes in high (HFI) and low (LFI) flows, annually and per season (DJF, MAM, JJA, SON) referring to North, Tropics and South latitude bands.

[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)
[Discussion Paper](#)


Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models

I. Giuntoli^{1,2,3}, J.-P. Vidal³, C. Prudhomme², and D. M. Hannah¹

¹School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK

²Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK

³Irstea, UR HHLY Hydrology-Hydraulics, Lyon, France

Correspondence to: I. Giuntoli (ignazio.g@gmail.com)

Abstract

Projections of changes in the hydrological cycle from Global Hydrological Models (GHMs) driven by Global Climate Models (GCMs) are critical for understanding future occurrence of hydrological extremes. However, uncertainties remain large and need to be better assessed. In particular, recent studies have pointed to a considerable contribution of GHMs that can equal or outweigh the contribution of GCMs to uncertainty in hydrological projections. Using 6 GHMs and 5 GCMs from the ISI-MIP multi-model ensemble, this study aims: (i) to assess future changes in the frequency of both high and low flows at the global scale using control and future (RCP8.5) simulations by the 2080s, and (ii) to quantify, for both ends of the runoff spectrum, GCMs and GHMs contributions to uncertainty using a 2-way ANOVA. Increases are found in high flows for northern latitudes and in low flows for several hotspots. Globally, the largest source of uncertainty is associated with GCMs, but GHMs are the greatest source in snow dominated regions. More specifically, results vary depending on the runoff metric, the temporal (annual and seasonal) and regional scale of analysis. For instance, uncertainty contribution from GHMs is higher for low flows than it is for high flows, partly owing to the different processes driving the onset of the two phenomena (e.g. the more direct effect of the GCMs precipitation variability on high flows). This study provides a comprehensive synthesis of where future hydrological extremes are projected to increase and where the ensemble spread is owed to either GCMs or GHMs. Finally, our results underline the ~~importance of using multiple GCMs and GHMs to envelope the overall uncertainty range and the~~ need for improvements in modeling snowmelt and runoff processes to project future hydrological extremes and the importance of using multiple GCMs and GHMs to envelope the uncertainty range provided by these two sources.

1 Introduction

The ongoing intensification of the water cycle at the global scale is expected to continue in the coming decades (Huntington, 2006; Stott et al., 2010). Projected changes in climate

variables from global General Climate Models (GCMs) indicate an increase in the frequency of hydrological extremes (Tebaldi et al., 2006; Seneviratne et al., 2012; Sillmann et al., 2013; Kharin et al., 2013). These hydrological shifts go hand in hand with a growing world population that will become ever more vulnerable with respect to access to water and food, and resilience to natural hazards (Lavell et al., 2012). In this context, global multi-model ensembles yield a valuable opportunity for climate projections and impact assessments. In hydrology, multi-model ensemble experiments – consisting of Global Hydrological Models (GHMs) fed by input forcing simulated by GCMs – can be used to project future changes in the water cycle and future hydrological extremes, using modeled variables such as precipitation, runoff, and soil moisture. In recent years, a number of studies have assessed the future changes in the global water cycle (e.g. Nohara et al., 2006; Hirabayashi et al., 2008; Sheffield and Wood, 2008). Although many of these studies have a representative number of GCMs in their ensembles, they rarely comprise more than one GHM, and this presents a limitation considering that GHMs provide more uncertainty than previously thought (Haddeland et al., 2011; Hagemann et al., 2013; Schewe et al., 2013; Prudhomme et al., 2014). In addition, the coarse temporal and spatial resolution of the climate signal used in these studies does not reflect well the potential changes in sub-monthly extreme events at the regional and local scale (Forzieri et al., 2014).

Recently, model inter-comparison projects like WaterMIP (Haddeland et al., 2011) and ISI-MIP (Warszawski et al., 2014) have allowed to include multiple GCMs and GHMs in global impact studies at unprecedented temporal (up to daily) and spatial (0.5°) resolution, thereby providing frameworks for consistent assessments of the terrestrial water cycle.

The ISI-MIP data set has been used to assess future changes in runoff at global and regional scales. Dankers et al. (2013) explored changes in 30 year return period of river flow showing that flood hazard is projected overall to increase globally, although not uniformly, and that decreases occur mainly in areas where the hydrograph is dominated by spring snowmelt. Schewe et al. (2013) assessed future water scarcity by analysing changes in mean annual runoff together with global population patterns, showing how the number of people living in water scarcity is projected to increase globally. Davie et al. (2013) inves-

5 tigated runoff changes across models by grouping GHMs into hydrological and biome (including CO₂ and vegetation dynamics) models, showing that while both types agree on the sign of runoff change for most regions of the world (with contrasting exceptions like West Africa where biome models moisten and hydrological models dry), models accounting for varying CO₂ yield more runoff than those with constant CO₂. Prudhomme et al. (2014) examined the future frequency of droughts using a -variable threshold method on daily runoff. They identified drought hotspots globally and observed, similarly to Davie et al. (2013), how biome models accounting for varying CO₂ concentrations tend to project more runoff with increasing CO₂ than the hydrological models. All of these studies emphasize how both
10 GCMs and GHM uncertainty contribute to the spread in projected changes in the hydrological cycle. Their findings highlight the importance of including different types of GHMs (~~as well as GCMs~~) in the comprehensive assessment and GCMs for making comprehensive assessments of uncertainty in climate impact studies.

In this context, modeling-induced uncertainty (i.e. inter-model spread of GCMs and
15 GHMs) has been expressed by looking at the variance across both types of models. For example, Schewe et al. (2013) and Dankers et al. (2013) used the ratio of the variances of GCM and GHMs results (for GCM: variance of the change across all GCMs for each GHM, and then averaged over all of the GHMs; and vice versa for GHMs). Similarly, using WaterMIP data, Hagemann et al. (2013) expressed the spread due to the choice
20 of model type using standard deviation of GCM and GHMs (for GCM: mean across all GHMs for each GCM, and standard deviation of the GCMs; and vice versa for GHMs). Prudhomme et al. (2014) omit the partition into GCM/GHM and express the uncertainty through the Signal-to-noise (by grouping results per type of model) in order to infer which global model type in the ensemble brings about highest agreement.

25 The studies cited above have provided useful knowledge on climate change impacts on the water cycle using the ISI-MIP dataset, however, a synthesis of future projections for high and low flows along with a consistent estimation of uncertainties is still missing. The present study builds on the work of Prudhomme et al. (2014) but shows on low flows of Prudhomme et al. (2014) , but introduces several new aspects. First, in addition to Firstly,

low flows (Q_{10}), ~~we now analyze high flows (Q_{95}) are now analysed~~ using an improved index extraction, ~~which uses a revisited version of the~~. The variable threshold method ~~for both flows with a shorter fixed time window (5day), overcoming some of the limitations of the 30day moving window~~ used in Prudhomme et al. (2014), has been revisited to overcome a limitation of the 30-day moving window for which gridcells were assigned lower threshold values than the theoretical threshold assigned (Q_{10}) (i.e. a tendency to capture fewer occurrences, an effect perhaps attributable to GHMs' slow emptying of reservoirs during the recession phase). A shorter 5-day fixed time window eliminates this effect. Note that, in order to gather further data for the estimate of the quantile flow, the period of analysis was increased from 30 to 34 years, starting four years earlier (1972 for control and 2066 for future). Secondly, we now analyse high flows (Q_{95}), with the same method used for low flows (5-d fixed-window variable threshold method). Dankers et al. (2013), who also analyzed high flows, have focused on a different metric (annual extreme monthly flood peak with 30-year return level), as their aim was to describe changes in flood hazard, while our focus is on change in frequency of high flow days. In our study high and low flows are hence identified jointly with the same ensemble of 5 GCMs and 6 GHMs. While comprising the same number of GCMs, the ensemble used by Prudhomme et al. (2014) uses one additional GHM (JULES) and Dankers et al. (2013) uses three additional GHMs (JULES, LPJmL, MATSIRO). We did not use these additional GHMs as they showed large areas with long pools of zero values hindering the index extraction, making them unsuitable for our analysis, especially for the low flows; additionally, JULES was run at a coarser resolution ($1.25\text{--}1.875^\circ$ vs. $0.5\text{--}0.5^\circ$) that would potentially influence the uncertainty analysis. Thirdly, we assess systematically the relative contribution of GHMs and GCMs to uncertainty using an analysis of variance (ANOVA) framework (as in e.g., Yip et al., 2011; Sansom et al., 2013). ~~With the exception of the JULES GHM.~~ This uncertainty assessment moves beyond the signal-to-noise ratio by Prudhomme et al. (2014), as the quantification of each source (GCM/GHM) to total uncertainty allows to describe the spatial variability of the contributions gridcell per gridcell. While (Dankers et al., 2013; Schewe et al., 2013) partition GCM/GIM uncertainty using ratios between the variances, our ANOVA approach adds the contribution

of the error (or residual) to the partition of the variance along with post-hoc testing on the residuals for model adequacy. We thus describe how high and low flows and inherent uncertainty vary at the seasonal and spatial scale, identifying areas where we have more confidence in the climate or in the hydrology (i.e. uncertainty is owed to GCMs or GHMs).
5 ~~Finally, we used the same dataset of GCMs and GHMs (i.e. 5GCMs and 6GHMs). JULES was left out because of the potential influence on the uncertainty analysis that could arise from its spatial grid mismatch (1.25–1.85vs.0.5–0.5). Thirdly,~~ to understand how the variance of the changes differs regionally, we carry out analysis at the regional scale expressing the ANOVA sum-of-squares of each source using homogeneous geo-climate
10 regions (Köppen–Geiger). This allows for an improved understanding of how the climate and hydrological processes drive uncertainty for both runoff ends.

By comparing an ensemble of GCMs (5) and GHMs (6) for future projections (2066–2099) against the historical period (1972–2005), this study aims (i) to assess future high and low flows changes at global and annual and seasonal scales, and (ii) to quantify the uncertainty
15 attributable to GHMs and GCMs using ANOVA. In the next section, the dataset and the different steps of the methodology are detailed. The results of projected hydrological extremes and respective uncertainty are presented in Sect. 3 before discussing the important and wider implications of this research in the fourth and final section.

2 Data and methods

20 The dataset used herein comes from the Inter-Sectorial Impact Model Intercomparison Project (ISI-MIP) (Warszawski et al., 2014) and consists of daily total un-routed runoff at a spatial resolution of 0.5 degrees from an ensemble of six GHMs forced with five CMIP5 GCMs bias-corrected climate (Hempel et al., 2013) for the historical (1972–2005) and future (2066–2099) periods under the RCP 8.5 scenario. The six GHMs are: H08, MPIHM, MacPDM, VIC, WBM, PcrGLOBWB (see Table A1 in the Appendix Appendix A for a summary of the main characteristics), and the five GCMs are: HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, NorESM1-M (refer to Warszawski et al. (2014) for
25

further details on the models and to www.isi-mip.org to access the simulation protocol). It should be noted that the selection of GHMs was dictated by temporal (daily runoff) resolution and time series tractability: models with lengthy pools of runoff equal to zero over large portions of the globe imposing constraints to the index extraction were not included (this aspect is described further in Appendix B). The selected model combinations form an ensemble of 30 experiments, each consisting of a historical and future period. none of the GHMs include varying CO₂.

Our analytical framework was composed of four steps: (i) time-series of days classified as high and low flows were extracted from daily total runoff record; (ii) high and low flows indices (i.e. change in frequency of high/low flows) were calculated (future-historical period) and mapped; (iii) ANOVA was carried out on the high and low flows indices considering GCMs and GHMs as factors; and (iv) the dominant uncertainty factors were explored for high and low flows across different climate regions based on the Köppen–Geiger classification.

To quantify high and low flows inter-annual variability, daily binary series (zero or one) were extracted for every land gridcell: high flows days, HFD; and low flows days, LFD. The series extraction uses daily varying threshold curves obtained from the daily runoff series for the historical period (1972–2005), which are then applied to historical period and future projections to identify days above (for HFD) or below thresholds (for LFD), as in e.g., for low flows, Sheffield and Wood (2008); Prudhomme et al. (2014). High flows are characterized by the 95th percentile (Q_{95} – runoff equaled or exceeded 5 % of the time) and low flows by the 10th percentile (Q_{10} – runoff equaled or exceeded 90 % of the time). For HFD, a ~~value of 0 (no value of 1~~ (high flow) ~~or 1 (high flows)~~ is assigned to each cell if the ~~runoff value of the cell~~ cell's runoff exceeds the Q_{95} value. ~~For LFD, a~~, otherwise a value of 0 (no ~~low flow~~) ~~or high flow~~ is assigned. For LFD, a value of 1 (low flows) is assigned to each cell if the cell's runoff is below the Q_{10} value. ~~A~~, otherwise a value of 0 (no low flow) is assigned. A comprehensive description of the threshold and binary series extraction together with an explanatory picture (Fig. B1) are provided in the Appendix. Gridcells showing little or no seasonal change in ~~daily runoff~~ the daily runoff of the control period (1972–2005) were

screened-out and represented in grey on the maps (for a comprehensive explanation of the masking see Appendix B). These screened-out gridcells are often located in arid or frozen regions where there is little or no runoff during long periods of the year and so the index extraction becomes intractable due to the presence of repeated zero values in the series.

We use indices to express the change in the frequency (in %) of: future high (HFI) and low (LFI) flows. These indices are calculated as follows: for each ensemble member HFI (LFI) is equal to the difference between the frequency (in %) of high (low) flows days ($100 \times \text{mean of HFD (LFD)}$) from the future (2066–2099) and historical period (1972–2005), for the whole year and per season (DJF and JJA). Both HFI and LFI are composed of 30 series (i.e. 6 GHMs fed by 5 GCMs each). The agreement in the change across ensemble members is expressed by the signal-to-noise ratio, S2N, calculated by dividing the median of the ensemble flow indices (HFI and LFI) by the inter-quartile range (75th percentile minus 25th percentile). The higher the S2N, the higher the members agreement in the signal, assuming signal greater than noise if $S2N > 1$.

In this study, the uncertainty is reflected by the spread of the flow indices due to the choice of GCM or GHM. To quantify the individual contribution of GCMs and GHMs to total uncertainty, a 2-factor ANOVA was carried out on the flow indices HFI and LFI for each gridcell. For this data set, model runs had no replicates, therefore the ANOVA model considers one case per treatment (Neter et al., 1999, ch. 21), so no interactions ($\alpha\beta_{ij} = 0$) and fixed factors levels ($n = 1$):

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (1)$$

where: Y_{ij} is the mean change for GCM_{*i*} and GHM_{*j*}, μ is a constant (the overall mean), α_i is the main effect for GCM at the *i*th level, β_j is the main effect for GHM at the *j*th level, ε_{ij} is the residual $\approx N(0, \sigma^2)$ iid. Thus, the variance is partitioned into two factors, GCMs and GHMs, plus the residuals. The results, expressed in terms of sum of squares, are used to quantify the factors contributions to the total variance, here considered as uncertainty as in e.g. Sansom et al. (2013). ANOVA models are reasonably robust against certain types of departures from the model (e.g. error terms not being exactly normally distributed).

Nonetheless, the suitability of the ANOVA model with the data at hand should be checked for serious departures from the conditions assumed by the model by looking at the residuals (Neter et al., 1999, ch. 18) and testing their normality (e.g. Lilliefors Test) and constancy of variance (e.g. Hartley Test). Unsatisfactory results would require remedial measures like data transformation or a modification of the model.

To understand how variance differs between climate regions, [the ANOVA sum of squares for all model combinations](#) are shown per Köppen–Geiger class ([Kottek et al., 2006](#)). [We used the Köppen–Geiger data classification based on present day proposed by Kottek et al. \(2006\) \(a link to the map is provided at the end of Table 1\).](#) A total of 15 (out of 31) regions are considered leaving out underrepresented regions with too few gridcells ($i.e.$ < 1000).

3 Results

Annual mean changes and associated S2N across all GHMs and GCMs are shown for HFI and LFI in Fig. 1a and b. For high and low flow indices, the mean changes vary spatially [and](#) in magnitude (Fig. 2) but they are positive generally, this means increases in number of days with (i) high flows, mostly over high northern latitudes; and (ii) low flows, spread over all latitudes with hotspots in: southern Europe; south western and mid Latin America; south eastern US and south eastern Canada; lower parts of Central Africa; north/north eastern China; and south western Australia. Regions screened-out represent 14 and 18 % of land for HFI and LFI, respectively. The S2N shows model agreement generally over the same regions for both indices (e.g., southern Europe, south western and mid Latin America, southeastern US). However, model agreement is found for HFI – but not for LFI – over Alaska, eastern Canada, and northwestern and eastern Russia. In some regions increases are not associated to a strong S2N (e.g. for high flows western China and the Horn of Africa). Mean changes and S2N for boreal winter (DJF) and summer (JJA), in Figs. 3 and 4 respectively, show an increased intensity with very similar spatial patterns to their annual counterparts in DJF for the high flows and in JJA for low flows. Conversely, high flows in

JJA show virtually no change, while low flows in DJF show decreases at high northern latitudes with high model agreement and increases elsewhere with smaller model agreement (S2N). This can be seen also in Fig. 2: the PDF (i.e. the density of the mean change percentage) stretches towards higher mean changes for high flows in DJF and for low flows in JJA. Global results are dominated by boreal seasonality (high flow changes dominant in DJF, and low flow changes dominant in JJA) as the majority of global land cells 65 % (of unmasked land) are located North of Latitude 23.5. The remainder of the land cells (35 %) are located within the Tropics and South latitude bands, and depict weak changes for high flows in all seasons, and increased changes for low flows in all seasons, though JJA's are more marked.

The results of the ANOVA across the 30 members of HFI and LFI are shown in Fig. 1c; they are expressed, for each factor, as the proportion of sum of squares divided by the total sum of squares (refer to Appendix C for residuals testing for model adequacy). For the high flows, the variance is explained mostly by the GCMs (yellow, 47 % of unmasked land, Fig. 1c), although the GHMs are the major factor over western Europe and central Canada (green, 28 % of unmasked land, Fig. 1c). For low flows, the proportions change: the GCMs (43 %) remain the major contributors over the globe, but GHMs (35 %) increase to a relative influence closer to the GCMs, and become the major factor in some northern (e.g. north eastern Russia) and southern (e.g. southern Africa, south western Australia) latitude regions. Seasonal results (Figs. 3c and 4c) are very similar to their annual counterparts in the case of high flows in DJF and low flows in JJA, whereas for high flows in JJA and for low flows in DJF higher residual rates (i.e. decreased overall GHM and GCM contributions) are found, perhaps owing to fewer events occurring in these seasons for both low and high flow indices.

To capture better the spatial distribution of the major sources of uncertainty, ANOVA results are aggregated by climatic homogeneous regions based on the climatological Köppen–Geiger classification. Scatterplots in Fig. 5 show the proportions of sums of squares of GHMs (y axis) vs. GCMs (x axis); medians for each climatic region are shown as their class letter and summarize the prominent factor of uncertainty. For both high and low flows

calculated over the year and seasonally, uncertainty in equatorial regions (A) is dominated by GCMs (median closest to the x axis); while in snow dominated climate (D) it is dominated by GHMs (median closest to the y axis). In warm temperate regions (C), uncertainty is slightly higher for GCMs than GHMs. In arid regions (B), the variance is not well explained by either GCMs or GHMs (median farthest from 1; i.e. residuals explain most of the variance), suggesting that reproducing hydroclimatology over these regions represents a challenge for both GCMs and GHMs. The ANOVA results for the whole year and those for winter and summer seasons (DJF and JJA shown in Figs. 3c and 4c) are quantified further in Table 1. This table provides a breakdown with both the regional and global results expressed for mean changes, S2N and percentage of sum of squares per factor at the annual and seasonal (DJF and JJA) scale. Looking jointly at the annual and seasonal results in Table 1, it is clear that the widespread dominance of the GCMs contribution to uncertainty is outweighed by the GHMs in the snow and ice dominated regions (D). This pattern is visible also on the scatter plots (Figs. 5 and 6) with the GHM uncertainty dominated regions (near y-axis) often populated by D regions for both HFI and LFI (although to a lesser extent for the former).

4 Discussion and conclusions

Using six global hydrological models (GHMs) fed by five global climate models (GCMs) under the RCP8.5 scenario, this study aimed to assess future high and low flows changes globally by 2080s, and to quantify the uncertainty attributable to GHMs and GCMs. We decided to focus solely on the uncertainty coming from GHMs and GCMs using as many ensemble members (from the ISI-MIP project dataset) as possible under the RCP8.5, in which change signals are expected to be larger (i.e. emissions continue to rise leading to global radiative forcing levels of 8.5 W m^{-2} by the end of the 21st century). The hydrological simulations used in this study do not account for anthropogenic influences (e.g. water abstraction, augmentation and artificial storage) or land-use changes.

High and low flows changes in the future (2066–2099) relative to the control period (1972–

2005) exhibit a number of robust large-scale features. Increases in high flow days were found at northern latitudes, with a strong signal over eastern Canada, Scandinavia, north-western Russia, and around the Bering Sea (i.e. eastern Russia and Alaska). Increases in low flow days were found in southern Europe, southwestern and central Latin America, southeastern USA, more southerly parts of Central Africa, and southwestern Australia. These patterns are largely consistent with the few other studies carried out on runoff at the global scale with several GHM-GCM combinations e.g. for high flows Dankers et al. (2013) and Hirabayashi et al. (2013), and low flows Van Huijgevoort et al. (2013) and Prudhomme et al. (2014) **specifically**; and for mean flows Davie et al. (2013), Schewe et al. (2013), and Hagemann et al. (2013). **However, in** More specifically, the comparison of flood hazard patterns by Dankers et al. (2013) with the changes in the occurrence of high flow days from our study, reveals some similarities, mostly northern North America and Northern Asia, while in some regions like North-eastern Europe patterns are opposite. Low flow patterns are similar to Prudhomme et al. (2014) although they find a weaker signal-to-noise. In this study we provide for the first time a comprehensive assessment of both ends of the runoff spectrum at the same time using the same dataset globally. Moreover, we undertake a consistent partition of uncertainty via ANOVA for both high and low flows, showing that GCMs provide the largest uncertainty, although the GHMs contribution can be substantial in particular regions. The results from our ANOVA framework are consistent with other global studies based on the ratios between the variances (or standard deviations) of ensemble members averaged per type of model (Dankers et al., 2013), Schewe et al. (2013), and Hagemann et al. (2013). In particular, uncertainty results, that Dankers et al. (2013) expressed with GCM/GHM variance, are in agreement with our findings for high flows in the southern hemisphere, mainly driven by GCM uncertainty, whereas there is less agreement for the northern hemisphere (in North America, Central Canada is GCM driven uncertainty, whereas it is GHM driven in our results). Uncertainty results for low flows from Prudhomme et al. (2014), expressed as S2N ratio, are not directly comparable, but as will be discussed later, the presence of the JULES GHM in their ensemble have shown to increase the uncertainty.

At the regional level, the uncertainty partition enables to delineate in which climate region each factor (GCMs or GHMs) provides the largest uncertainty at the annual and seasonal scales. Notably, for snow and ice dominated, polar regions, and arid zones, GHMs bring about the largest portion of uncertainty, especially for low flows. This is likely to reflect uncertainty in the way the hydrological storage-release processes can modify the climate signal, particularly where these storage components are relatively large or water residence times high – hence the importance of considering several GHMs in studying changes in high and low flows. GCMs and GHMs uncertainty shares are similar for HFI and LFI globally, although the spatial patterns differ slightly (e.g., northeastern Russia and southwestern Australia, and Alaska are GCM driven in HFI, and GHM driven in LFI). This could reflect different dominant processes for high and low flows generation, with high flows events mainly driven by precipitation inputs or snow/ice-melt (i.e. atmospheric-driven processes); whereas low flows events develop over longer durations and are influenced more by land-surface processes like evaporation, infiltration and storage, which are simulated by the GHMs, each one with its own scheme and parametrization: e.g. for evapotranspiration, Penman–Monteith, Hamon (Haddeland et al., 2011 and Table A1 in the Appendix). Haddeland et al. (2011) have identified in the snow scheme employed by different GHMs a major source of difference between the model runoff simulations, and recent studies, at global (e.g. Hagemann et al., 2013) and regional scale (e.g. Jung et al., 2012) hint to an increase in uncertainty in snow dominated regions. Our study shows that in snow dominated and arid regions GHMs uncertainty equals or outweighs GCMs uncertainty for both high and low flows, highlighting the importance of comprising balanced sets of both global hydrological and climate models to envelope the overall uncertainty in these regions.

To put the current study in context and to provide suggestions for further studies, it is worth making a few considerations on the hydrological index extraction and clarify a few aspects of the uncertainty partition concerning the method and the data set we used.

The identification of high and low flows over long time series, and particularly over climate projections, is non trivial. As an illustration, van Huijgevoort et al. (2014) in their multi-model ensemble study on droughts, report that applying the threshold level method to the future

period using a threshold derived from the control period can lead to spurious pooling of drought events. They suggest that future changes could be accounted for by linking the drought threshold to adaptation scenarios like Vidal et al. (2012) did over France. Wanders et al. (2014) used a transient threshold level method for a moving reference period, in order to reflect the changes in hydrological regime over time, finding that the non-transient threshold method projected larger shares of areas in drought (except in snow dominated regions). For our study, the threshold was calculated over the control period, as changes in future extremes with respect to present day were sought. In general, the selection of threshold approach should consider that if, on the one hand, a consistent pooling of extreme events may be hampered by incremental shifts or shape changes of the hydrograph throughout the future, on the other hand, when assessing the changes in frequency with respect to the present, information on the present used for comparison is lost when the threshold adapts throughout the projections.

The model runs used in this study have no replicates; therefore, our ANOVA partition set-up poses some limitations as it assumes that the factors do not interact (no degrees of freedom are available for the estimation of the experimental error). However, interactions between the factors may indeed be present and, as pointed out by (Bosshard et al., 2013), these interactions may represent uncertainty contributions that do not behave linearly: e.g. a snowmelt bias of a GHM may depend on the temperature projection of the driving GCM that could lead to a nonlinear response in the simulated runoff. This could in part explain the high rate of residuals' contribution seen in some gridcells for which potential interactions hinder the ANOVA to properly disclose the factors main effects. To avoid this drawback multiple model runs would be necessary.

Bias correction and CO₂ and vegetation dynamics represent other sources of uncertainty that were not accounted for in this study, though their influence should be further investigated in future works. Bias-correction is commonly used to overcome bias inconsistencies between GCMs and impact models (i.e. GHMs) in climate impact studies; however, this technique alters the model output by e.g. reducing the inter-GCM variability and potentially their contribution to total uncertainty in climate projections (Dankers et al., 2013; Wada

et al., 2013), and it is argued that its use is not always justified (Ehret et al., 2012). Hagemann et al. (2011) even found that uncertainty due to bias-correction can be of the same order of magnitude as that related to the choice of GCM or GHM. As Huber et al. (2014) points out, findings on relative contributions of GCMs and GHMs to total impact uncertainty would need to stand the test of using non bias-corrected runs, but runs that have not been bias-corrected (with a method designed to preserve the long term trends in temperature and precipitation projections, Hempel et al., 2013) are unavailable within ISI-MIP or with the same GCM/GHM combinations.

As mentioned in the introduction, biome models have shown a larger spread than GHMs without varying CO₂ and vegetation dynamics processes, and it is argued that, due to the additional processes that they simulate, the inclusion of biome models in multi-model ensemble studies is important to capture ~~all~~ a comprehensive range of uncertainty (Davie et al., 2013; Prudhomme et al., 2014). Within our study specifically, biome models with runs at daily resolution were JULES and LPJmL. These models were excluded primarily for intractability in low flows analysis, and for grid resolution mismatch (JULES) ~~and for intractability in low flows analysis~~. Therefore, uncertainty from varying CO₂ is not sampled and could suggest overconfidence (or bias) in favor of non biome GHMs, which simulate less runoff than biome models. During our exploratory analysis we actually included JULES in the ensemble and found that the uncertainty was driven towards the GHM source (in agreement with Prudhomme et al. (2014) who found higher S2N (stronger agreement between the models, i.e. the decreased uncertainty) when considering the ensemble without JULES. However, the inclusion of models in the ensemble must be compatible with the applicability of the method and the biome models available through ISI-MIP proved to hamper the global comparison assessment for the heavy land masking resulting from the presence of large pools of zero values in their time series. As shown in Table A2, low flows index extraction was vetoed over large areas of the globe, ultimately leaving 61% and 20% of land cells for JULES and LPJmL respectively (note that the masking is formed by superimposing masking from each GHM-GCM combinations). Also, the number of gridcells representing global land have a different order of magnitude for JULES compared to the

other GHMs (7558 and 67420 respectively, i.e. a ratio of 1 cells to 9), and the effect of different resolution cannot be assessed properly as runs with the same setup at lower resolutions were not available. Index extraction for high flows proved more favorable, but we adopted the pragmatic approach of using the largest possible ensemble of models common to both high and low flows. We are aware that the inclusion of multiple models is not sufficient to fully scope model uncertainty due to resolution and structural errors that are common across models and place a limit to the confidence we obtain from robustness (Knutti, 2010). However, our results demonstrated that, even excluding biome models and other model structure differences in the ISI-MIP ensemble, large uncertainty in the signal of changes in high and low flows is attributable to GHMs and not only on GCMs.

Were biome models shortcomings not present, their inclusion in our ensemble would have required a modification of our uncertainty partition strategy because the presence of outliers (likely introduced by biome models) would limit our ANOVA model (whose assumptions include no or minimal presence of outliers). For their distinct behavior from the other GHMs, biome models could be considered as a factor level in a 2-way ANOVA framework with unequal sample sizes (Neter et al., 1999, ch. 23), i.e. the spread of future hydrological extremes would be examined as the function of factor 1 – the type of hydrological model (level 1: 6 GHMs; level 2: 2 biome models) and factor 2 – the GCMs.

Finally, the focus of our uncertainty analysis was on GCMs and GHMs, therefore the effect of emission scenarios (RCPs) was neglected. The few studies that have considered this aspect hint at a relatively small role of emission scenarios (Hagemann et al., 2013; Wada et al., 2013) all throughout the 21st century when compared to GCMs and GHMs, which play a stronger role in uncertainty contribution over most of the globe.

To conclude, knowledge of the dominant source of uncertainty in climate-to-hydrology signal is critical to modelers for improving modeling of the terrestrial water cycle and to scientists for putting together targeted multi-model ensembles for climate impact studies. In addition to GHMs and GCMs, further work is needed to assess the degree to which internal variability, bias correction, biome models (i.e. GHMs that simulate vegetation dynamics and varying CO₂), and emission scenarios contribute to total uncertainty.

Appendix A: Global hydrological models

The Global Hydrological Models (GHMs) vary in the types of processes represented and the parameterizations used. Table A1 summarizes the main processes included in the GHMs used in this study. Input variables are listed under ?Meteorological Forcings? they include: Surface air temperatures; Precipitation; Surface radiation; Near-surface windspeed; Surface air pressure; Near-surface relative humidity. Except the last one, as reported in the ISI-MIP Protocol, all of these variables consist of bias?corrected climate data from the GCMs participating in the CMIP5 and cover the time period from 1950 to 2099 (1950-1970 are usually used for spin-up). All variables have daily and monthly frequency. Fig. B2 shows, for the control period, inter-annual dynamics in mean daily runoff simulated by the GHMs (in row) for the different GCMs for selected representative gridcells, one per main Köppen–Geiger region (A-Tropical, B-Arid, C-Temperate, D-Cold, E-Polar).

Appendix B: High and low flows binary series extraction and masking

The schematic of extraction of binary series of days under high (HFD) and low (LFD) flows is shown in Fig. B1. The threshold curves are obtained by linearly interpolating percentiles calculated over fixed 5 day windows (e.g. 1–5 December, 6–10 December, and so forth, i.e. 73 for the whole year) of the historical period runoff (i.e. December 1971 to December 2005), having considered the hydrological year from December to November.

The percentiles are Q_{95} (runoff equaled or exceeded 5 % of the time) for HFD, and Q_{10} (runoff equaled or exceeded 90 % of the time) for LFD. In general, the identification of high and low flows at the global scale imposes the selection of a universal threshold level serving many hydrological regimes and climate regions at once (thereby pooling events that may not always be extreme) and it is based on physical processes: low flows are generally characterized by a slower onset, and a longer duration, while and high flows by a sudden onset, and a shorter duration. Accordingly, high and low flows are not necessarily symmetric with respect to the median flow (Q_{50}). For low flows in particular, the choice of

Q₁₀ comes from seeking a sufficiently low quantile without compromising the analysis, as quantiles lower than 10 % become intractable for the large presence of zero pools in some time series. This is in agreement with e.g. citepGudmundsson2012 who showed how the performance of a similar set of WaterMIP global models decreased systematically from high Q₉₅ to low Q₅ runoff percentile over Europe.

The choice of a fixed 5 day time window with interpolation was preferred over the 30 day moving average used in e.g. Prudhomme et al. (2014) because the latter had shown some limitations with regards to the low flows quantile extraction. The effect of leveling out over 30 days could lead to lower values than expected in the control period (10 % by design). In addition, we wanted to use the same framework for high and low flows and considered 5 day to be appropriate to identify both types of events. The choice of a linear interpolation was preferred over the moving window approach to minimize dependence (i.e. inertia) within quantile estimates with the following rationale: (i) Moving average aims to smooth out wiggles for a less spiky identification of hydrological events like droughts that could result into erratic threshold crossings, thereby pooling several times over the same event; however, its quantile estimates use the same information from neighboring days (as many as the time window), resulting in a quantile series holding a correlation that is higher the longer the time window, potentially leading to inadvertent effects of large inertia during the extraction of the hydrological index. (ii) In our case, as we count high (low) flow days (as opposed to single events), smoothing the threshold is unnecessary. (iii) A 1 day window would assure a series of independent quantile estimates, but the computation over 34 points (i.e. 34 years of the control period) was considered insufficient for quantile estimation. (iv) Seeking a representative number of points for quantile extraction (170, i.e. 5 days \times 34 years), we decided to compute the quantile by extracting a point every 5 days and extrapolating values for intermediate days to the next 5 days point; as a result threshold values were obtained with a non recursive use of data, thereby minimizing dependence.

The index extraction described above is not applicable when the runoff is very low, i.e. when long periods of the year have the same value. Gridcells-Therefore, with reference to the control period (1972-2005), gridcells showing little or no seasonal change in daily runoff

were screened out (represented in grey on the maps) ~~on the basis of the~~ using the 5 day percentiles series that form the threshold curves (i.e., one mask for HF and one for LF) following these rules: (i) percentiles are equal to zero for more than one third of the year (ii) standard deviation of percentiles of first and/or second half year equals zero (iii) annual percentiles Q_{10} and Q_{95} series are equal.

Table A2 shows percentages of available land gridcells after screening for the different GCM-GHM combinations and runoff percentile. Although screened gridcells could become seasonal through the climate projection - e.g., Alessandri et al. (2014) investigated the expansion and retreat of specific climate boundaries (Mediterranean climate in Europe and western USA) using CMIP5 data - we neglect this aspect as our base reference for changes in projections is the control period.

Mean changes for each ensemble member (GHM-GCM combination) are shown in Fig. B3 and Fig. B4, for high and low flows respectively.

Appendix C: Tests on ANOVA's residuals

To verify whether the ANOVA model assumptions hold, statistical tests were performed on the ANOVA's residuals. For every unmasked gridcell, for both HFI and LFI, residuals were assessed as follows: we tested (i) normality with the Lilliefors Test; and then, for gridcells for which the null hypothesis (that the residuals' vector comes from a distribution in the normal family) was not rejected, we tested (ii) constancy of variance with the Hartley Test. Results for the annual and seasonal ANOVAs show that HFI has higher rates of residuals for which the hypotheses of normality and constancy of variance were rejected compared to the LFI. For the year, the percentages of unmasked gridcells not meeting the residuals requirements were: HFI 22 % not normal, 15 % no constant variance, for a total of 37 % globally; LFI 12 % not normal, 15 % no constant variance, for a total of 27 % globally. JJA and DJF have the lowest proportions of residuals requirements not met for HFI and LFI respectively. We also applied the ANOVA on HFI and LFI transformed via the normal-score method (seeking normality of the data); this showed lower percentages of cells not satisfying the

ANOVA assumptions of normality and constant variance (HFI: 7.5 and 11%; and LFI: 7 and 12% respectively) for a total of 19% globally. It should be noted that the residuals' contribution to uncertainty tends to be lower for the transformed data (e.g. gridcells with residuals' dominated uncertainty decreased by 6% for HFI and 1% for LFI). Because the partition of uncertainty between GCMs and GHMs are similar from both ANOVA applied to raw and transformed data sets, and because the areas of non-satisfaction of normality are not located where the residuals dominate the uncertainty, we discussed results obtained from the raw, non transformed data.

Acknowledgements. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling responsible for the Coupled Model Intercomparison Project (CMIP); and we thank the CMIP climate and hydrology modeling groups for their model outputs. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provided co-ordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This work has been conducted under the framework of the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP). The ISI-MIP Fast Track project was funded by the German Ministry of Education and Research, with project funding reference number 01LS1201A. To obtain access to data, please refer to the "data archive" section of the project portal www.isi-mip.org. I. Giuntoli was funded by a PhD scholarship from the UK Natural Environment Research Council (NE/YXS1270382) to the University of Birmingham in collaboration with the Centre for Ecology and Hydrology, Wallingford. acknowledgements

References

- Alessandri, A., De Felice, M., Zeng, N., Mariotti, A., Pan, Y., Cherchi, A., Lee, J.-Y., Wang, B., Ha, K.-J., Ruti, P., and Artale, V.: Robust assessment of the expansion and retreat of Mediterranean climate in the 21st century, *Sci. Rep.*, 4, <http://dx.doi.org/10.1038/srep07211>, 2014.
- Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., and Schär, C.: Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, *Water Resources Research*, 49, 1523–1536, doi:10.1029/2011WR011533, 2013.
- Dankers, R., Arnell, N. W., Clark, D. B., Falloon, P. D., Fekete, B. M., Gosling, S. N., Heinke, J., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: First look at changes in flood

hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble, *Proceedings of the National Academy of Sciences*, pp. 1–5, 2013.

Davie, J. C. S., Falloon, P. D., Kahana, R., Dankers, R., Betts, R., Portmann, F. T., Clark, D. B., Itoh, A., Masaki, Y., Nishina, K., Fekete, B., Tessler, Z., Liu, X., Tang, Q., Hagemann, S., Stacke, T., Pavlick, R., Schaphoff, S., Gosling, S. N., Franssen, W., and Arnell, N.: Comparing projections of future changes in runoff and water resources from hydrological and ecosystem models in ISI-MIP, *Earth System Dynamics Discussions*, 4, 279–315, <http://www.earth-syst-dynam-discuss.net/4/279/2013/>, 2013.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", *Hydrology and Earth System Sciences*, 16, 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F., and Bianchi, A.: Ensemble projections of future streamflow droughts in Europe, *Hydrology and Earth System Sciences*, 18, 85–108, 2014.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, *Journal of Hydrometeorology*, 12, 869–884, 2011.

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C.: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models, *Journal of Hydrometeorology*, 12, 556–578, 2011.

Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F., and Wiltshire, a. J.: Climate change impact on available water resources obtained using multiple global climate and hydrology models, *Earth System Dynamics*, 4, 129–144, 2013.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, *Earth System Dynamics*, 4, 219–236, 2013.

Hirabayashi, Y., Kanai, S., Emori, S., Oki, T., and Kimoto, M.: Global projections of changing risks of floods and droughts in a changing climate, *Hydrological Sciences Journal*, 53, 754–772, 2008.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanai, S.: Global flood risk under climate change, *Nature Climate Change*, 3, 816–821, 2013.

- Huber, V., Schellnhuber, H. J., Arnell, N. W., Frieler, K., Friend, A. D., Gerten, D., Haddeland, I., Kabat, P., Lotze-Campen, H., Lucht, W., Parry, M., Piontek, F., Rosenzweig, C., Schewe, J., and Warszawski, L.: Climate impacts research: beyond patchwork, *Earth System Dynamics Discussions*, 5, 721–746, 2014.
- 5 Huntington, T. G.: Evidence for intensification of the global water cycle: Review and synthesis, *Journal of Hydrology*, 319, 83–95, 2006.
- Jung, I.-W., Moradkhani, H., and Chang, H.: Uncertainty assessment of climate change impacts for hydrologically distinct river basins, *Journal of Hydrology*, 466–467, 73–87, doi:10.1016/j.jhydrol.2012.08.002, 2012.
- 10 Kharin, V., Zwiers, F., Zhang, X., and Wehner, M.: Changes in temperature and precipitation extremes in the CMIP5 ensemble, *Climatic Change*, 119, 345–357, 2013.
- Knutti, R.: The end of model democracy?, *Climatic Change*, 102, 395–404, doi:10.1007/s10584-010-9800-2, 2010.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, *Meteorologische Zeitschrift*, 15, 259–263, 2006.
- 15 Lavell, A., Oppenheimer, M., Diop, C., Hess, J., Lempert, R., Li, J., Muir-Wood, R., and Myeong, S.: Climate change: new dimensions in disaster risk, exposure, vulnerability, and resilience., Cambridge University Press, Cambridge, UK, and New York, NY, USA, 2012.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W.: *Applied Linear Statistical Models*, vol. 1, McGraw-Hill, 1999.
- 20 Nohara, D., Kitoh, A., Hosaka, M., and Oki, T.: Impact of Climate Change on River Discharge Projected by Multimodel Ensemble, *Journal of Hydrometeorology*, 7, 1076–1089, 2006.
- Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment., *Proceedings of the National Academy of Sciences of the United States of America*, 111, 3262–7, 2014.
- Sansom, P. G., Stephenson, D. B., Ferro, C. a. T., Zappa, G., and Shaffrey, L.: Simple Uncertainty Frameworks for Selecting Weighting Schemes and Interpreting Multimodel Ensemble Climate Change Experiments, *Journal of Climate*, 26, 4017–4037, 2013.
- 30 Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F.,

- Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, *Proceedings of the National Academy of Sciences*, 2013.
- Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment, Cambridge University Press, Cambridge, UK, and New York, NY, USA, 2012.
- Sheffield, J. and Wood, E. F.: Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations, *Climate Dynamics*, 31, 79–105, 2008.
- Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections, *Journal of Geophysical Research: Atmospheres*, 118, 2473–2493, 2013.
- Stott, P. A., Gillett, N. P., Hegerl, G. C., Karoly, D. J., Stone, D. A., Zhang, X., and Zwiers, F.: Detection and attribution of climate change: a regional perspective, *Wiley Interdisciplinary Reviews: Climate Change*, 1, 192–211, 2010.
- Tebaldi, C., Hayhoe, K., Arblaster, J. M., and Meehl, G. A.: Going to the Extremes, *Climatic Change*, 79, 185–211, 2006.
- van Huijgevoort, M., van Lanen, H., Teuling, A., and Uijlenhoet, R.: Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge, *Journal of Hydrology*, 512, 421–434, doi:<http://dx.doi.org/10.1016/j.jhydrol.2014.02.060>, 2014.
- Van Huijgevoort, M. H. J., Hazenberg, P., Van Lanen, H. a. J., Teuling, a. J., Clark, D. B., Folwell, S., Gosling, S. N., Hanasaki, N., Heinke, J., Koirala, S., Stacke, T., Voss, F., Sheffield, J., and Uijlenhoet, R.: Global Multimodel Analysis of Drought in Runoff for the Second Half of the Twentieth Century, *Journal of Hydrometeorology*, 14, 1535–1552, 2013.
- Vidal, J.-P., Martin, E., Kitova, N., Najac, J., and Soubeyroux, J.-M.: Evolution of spatio-temporal drought characteristics: validation, projections and effect of adaptation scenarios, *Hydrology and Earth System Sciences*, 16, 2935–2955, doi:[10.5194/hess-16-2935-2012](https://doi.org/10.5194/hess-16-2935-2012), 2012.
- Wada, Y., Wisser, D., Eisner, S., Flörke, M., Gerten, D., Haddeland, I., Hanasaki, N., Masaki, Y., Portmann, F. T., Stacke, T., Tessler, Z., and Schewe, J.: Multimodel projections and uncertainties of irrigation water demand under climate change, *Geophysical Research Letters*, 40, 4626–4632, doi:[10.1002/grl.50686](https://doi.org/10.1002/grl.50686), 2013.

Wanders, N., Wada, Y., and Van Lanen, H. A. J.: Global hydrological droughts in the 21st century under a changing hydrological regime, *Earth System Dynamics Discussions*, 5, 649–681, doi:10.5194/esdd-5-649-2014, 2014.

5 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): project framework., *Proceedings of the National Academy of Sciences of the United States of America*, 111, 3228–32, 2014.

Yip, S., Ferro, C. a. T., Stephenson, D. B., and Hawkins, E.: A Simple, Coherent Framework for Partitioning Uncertainty in Climate Predictions, *Journal of Climate*, 24, 4634–4643, 2011.

Table 1. Summary of mean changes, signal-to-noise S2N, and sources of variance for high and low flows at the annual and seasonal (DJF, JJA) scale, at the global and climate region scale. The first source of variance is shown in bold, the second one in italic font.

Köppen-Geiger Class*	Area [km²]	YEAR			DJF			JJA			Source of variance			Source of variance		
		Mean ¹ change [%]	Signal ² to noise adim.	Source of variance GCM [%]	Signal to noise adim.	Mean change [%]	Signal to noise adim.	Source of variance GCM [%]	Signal to noise adim.	Mean change [%]	Signal to noise adim.	Source of variance GCM [%]	Signal to noise adim.	Source of variance GCM [%]	Signal to noise adim.	Source of variance GCM [%]
HIGH FLOWS																
Equatorial																
1 Af	2468	3.1	0.5	65.5	14.2	20.3	3.4	0.5	64.4	11.4	24.3	3.2	0.6	64.0	12.4	23.6
2 Am	1836	1.4	0.5	61.6	15.5	22.9	2.8	0.4	60.2	14.8	25.0	0.7	0.9	65.9	11.0	23.1
3 Aw	6017	1.6	0.6	52.7	20.4	27.0	2.9	0.4	56.3	18.3	25.5	0.5	0.9	47.0	19.7	33.3
Arid																
4 Bwk	3095	4.1	0.5	49.7	19.8	30.5	5.5	0.4	47.1	20.7	32.3	2.3	0.7	42.9	20.3	36.8
5 Bwh	3139	1.2	0.6	39.3	28.5	32.2	3.1	0.6	35.6	29.6	34.8	0.1	0.7	37.7	25.3	37.1
6 BSk	4255	6.4	0.4	40.5	25.1	34.4	7.0	0.4	39.6	23.9	36.4	5.1	0.5	36.2	25.5	38.3
Warm temp.																
7 Cfa	2955	-0.6	0.6	56.1	26.8	17.1	-0.5	0.4	57.1	22.9	20.0	-0.8	0.8	55.4	20.1	24.5
8 Cfb	2360	0.4	0.9	45.0	35.9	19.0	1.8	0.6	40.5	37.7	21.9	-0.5	1.0	45.4	29.1	25.5
9 Csa	1099	-1.2	1.5	45.3	32.8	21.9	-0.8	1.3	50.8	30.2	19.0	-2.0	1.9	31.0	30.5	38.5
10 Cwa	1504	0.9	0.6	43.8	28.0	28.2	0.9	0.4	45.6	27.6	26.8	1.1	0.8	49.8	23.9	26.3
Snow																
11 Dfb	4459	5.4	0.7	42.7	37.1	20.2	17.5	0.7	38.2	40.3	21.5	-1.1	0.8	57.3	21.6	21.1
12 Dfc	11008	18.6	1.2	38.0	38.8	23.2	37.5	1.0	36.0	39.7	24.2	0.2	0.7	39.7	34.2	26.2
13 Dfd	1405	27.4	1.1	42.7	22.4	34.9	39.8	0.9	38.1	26.2	35.6	7.4	0.4	13.6	55.6	30.8
14 Dwb	1311	6.7	0.5	29.1	46.1	24.8	11.0	0.4	25.7	41.9	32.3	1.4	0.3	35.0	41.9	23.1
Polar																
15 ET	5937	26.3	1.3	42.9	36.1	20.9	40.6	1.2	43.7	32.2	24.1	5.5	0.5	34.3	41.0	24.7
Global	128.9M	6.5	0.7	46.5	28.0	25.5	11.8	0.6	45.5	27.5	27	1.3	0.8	45.0	25.4	29.7
LOW FLOWS																
Equatorial																
1 Af	2463	14.6	0.5	58.7	18.1	23.1	12.7	0.3	56.5	20.4	23.1	15.4	0.5	50.1	17.1	32.8
2 Am	1834	23.7	0.7	57.0	25.2	17.8	19.7	0.4	50.5	29.1	20.4	27.1	0.7	59.5	18.1	22.4
3 Aw	5997	21.7	0.7	52.4	28.5	19.0	18.4	0.5	48.2	29.8	22.0	25.7	0.6	50.8	28.3	20.9
Arid																
4 Bwk	2927	15.6	0.5	41.4	31.0	27.6	14.3	0.5	40.0	31.1	28.9	17.1	0.4	38.4	32.1	29.5
5 Bwh	2821	20.2	0.6	32.1	42.6	25.3	18.4	0.5	29.2	42.8	28.0	22.2	0.6	30.3	42.9	26.7
6 BSk	2693	14.3	0.6	35.9	32.1	32.0	13.5	0.6	35.0	31.2	33.7	15.1	0.5	33.8	33.3	33.0
Warm temp.																
7 Cfa	2950	18.2	1.0	49.1	32.9	18.1	17.6	0.7	47.5	32.4	20.0	19.1	0.9	44.1	30.9	25.0
8 Cfb	2358	20.2	1.1	51.7	32.3	16.0	15.2	0.8	43.7	36.4	19.9	24.4	1.0	46.2	33.6	20.1
9 Csa	1096	35.7	1.4	47.0	37.6	15.5	31.0	1.3	48.7	35.8	15.5	41.9	1.4	41.0	37.7	21.3
10 Cwa	1500	18.5	0.8	42.0	39.4	18.5	18.1	0.7	39.5	39.8	20.7	18.4	0.7	44.7	34.1	21.2
Snow																
11 Dfb	4440	15.8	0.8	50.6	28.1	21.3	4.1	0.5	29.8	43.5	26.7	26.3	0.9	52.4	26.0	21.6
12 Dfc	10920	8.7	0.5	33.6	44.8	21.7	-2.0	1.5	17.4	45.1	37.5	25.0	0.8	38.9	43.1	18.1
13 Dfd	1402	-2.5	0.7	15.3	59.3	25.4	-5.7	2.3	16.8	40.1	43.1	4.4	0.2	14.5	66.4	19.1
14 Dwb	1306	9.5	0.3	26.8	48.5	24.7	9.9	0.3	23.3	47.3	29.4	11.4	0.5	31.7	46.7	21.5
Polar																
15 ET	5650	3.4	0.5	29.8	45.0	25.2	-1.7	2.1	20.2	37.9	41.9	14.3	0.5	35.2	46.4	18.3
Global	122M	16.1	0.7	43.1	34.8	22.1	11.8	0.8	36.6	35.9	27.6	21.5	0.7	42.5	34.2	23.3

1st, 2nd Source of variance.

* Mean change weighted over gridcells surface areas. * Signal-to-noise weighted over gridcells surface areas.

* The map can be downloaded at: http://koeppen-geiger.vu-wien.ac.at/pdf/kottek_et_al_2008_A4.pdf.

Table A1. Global Hydrological Models main characteristics (after Prudhomme et al., 2014).

Model name ^a	Time step	Meteorological forcing ^b	Energy balance	Evaporation scheme	Runoff scheme	Snow scheme
H08	Daily	$R, S, T, W, Q, LW, SW, SP$	Yes	Bulk formula	Saturation excess, nonlinear	Energy balance
MPI-HM	Daily	P, T, W, Q, LW, SW, SP	No	Penman–Monteith	Saturation excess, nonlinear	Degree-day
Mac-PDM.09	Daily	P, T, W, Q, LWn, SW, SP	No	Penman–Monteith	Saturation excess, nonlinear	Degree-day
VIC	Daily, ³ h snow	$P, Tmax, Tmin, W, RH, LW, SW, SP$	Snow only	Penman–Monteith	Saturation excess, nonlinear	Energy balance
WBM	Daily	P, T	No	Hamon	Saturation excess	Empirical temp and precip based formula
PCRGlobWB	Daily	P, T	No	Hamon	Infiltration excess, saturation excess, groundwater	Degree-day

^a All of the 6 models were run at the spatial resolution of $0.5^\circ \times 0.5^\circ$.

^b LW: downwelling longwave radiation; LWn: net longwave radiation; P : precipitation rate (rain and snow calculated in the model); Q : air specific humidity; R : rainfall rate; RH: relative humidity; S : snowfall rate; SP : surface pressure; SW : downwelling shortwave radiation; T : air temperature; Tmax(min): daily maximum (minimum) air temperature; W : wind speed.

Table A2. Percentage of available land gridcells after masking per GHM-GCM model combination.

			GCM				
			HadGEM	IPSL	MIROC	GFDL	NorESM
GHM	H08	<u>Q₁₀</u>	<u>99.97</u>	<u>99.82</u>	<u>99.96</u>	<u>99.96</u>	<u>99.95</u>
		<u>Q₉₅</u>	<u>99.98</u>	<u>99.98</u>	<u>99.98</u>	<u>99.99</u>	<u>99.98</u>
	MPIHM	<u>Q₁₀</u>	<u>89.85</u>	<u>89.14</u>	<u>89.69</u>	<u>89.68</u>	<u>89.68</u>
		<u>Q₉₅</u>	<u>92.75</u>	<u>92.24</u>	<u>92.52</u>	<u>93.17</u>	<u>93.08</u>
	MacPDM	<u>Q₁₀</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>
		<u>Q₉₅</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>
	VIC	<u>Q₁₀</u>	<u>96.25</u>	<u>96.25</u>	<u>96.47</u>	<u>96.59</u>	<u>96.39</u>
		<u>Q₉₅</u>	<u>99.48</u>	<u>97.72</u>	<u>99.41</u>	<u>99.20</u>	<u>99.36</u>
	WBM	<u>Q₁₀</u>	<u>96.19</u>	<u>96.29</u>	<u>95.72</u>	<u>96.02</u>	<u>96.27</u>
		<u>Q₉₅</u>	<u>97.38</u>	<u>97.97</u>	<u>96.81</u>	<u>97.75</u>	<u>97.58</u>
	PCRGLOBWB	<u>Q₁₀</u>	<u>90.91</u>	<u>91.17</u>	<u>90.39</u>	<u>91.26</u>	<u>90.71</u>
		<u>Q₉₅</u>	<u>92.92</u>	<u>92.84</u>	<u>92.16</u>	<u>93.16</u>	<u>92.79</u>
	JULES*	<u>Q₁₀</u>	<u>64.07</u>	<u>64.05</u>	<u>65.45</u>	<u>66.06</u>	<u>66.59</u>
		<u>Q₉₅</u>	<u>84.71</u>	<u>89.16</u>	<u>91.39</u>	<u>89.57</u>	<u>91.06</u>
	LPJmL*	<u>Q₁₀</u>	<u>26.97</u>	<u>25.07</u>	<u>25.95</u>	<u>26.12</u>	<u>26.89</u>
		<u>Q₉₅</u>	<u>70.22</u>	<u>67.27</u>	<u>69.76</u>	<u>68.50</u>	<u>69.72</u>
	MATSIRO*	<u>Q₁₀</u>	<u>25.73</u>	<u>23.27</u>	<u>29.60</u>	<u>25.39</u>	<u>27.70</u>
		<u>Q₉₅</u>	<u>64.56</u>	<u>61.26</u>	<u>67.15</u>	<u>69.10</u>	<u>67.42</u>

* Models not included in the ensemble.

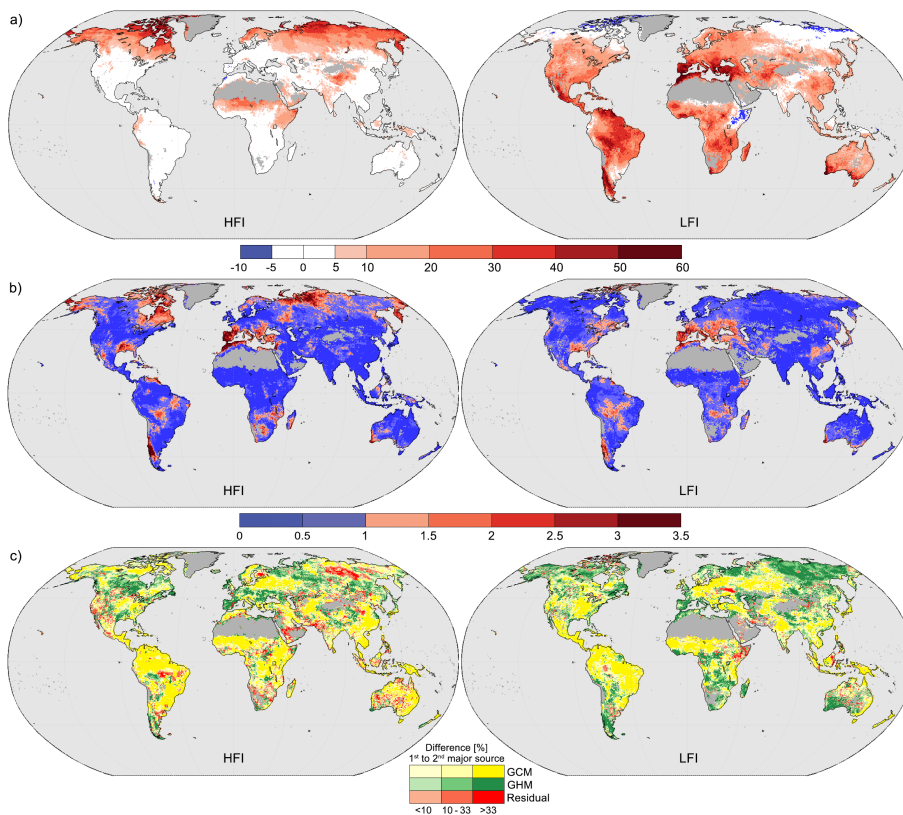


Figure 1. Change in the frequency (in %) of days under high (left) and low (right) flows conditions for the period 2066–2099 relative to 1972–2005, based on a multi-model ensemble MME experiment under RCP8.5 from five Global Climate Models and six Global Hydrological Models: **(a)** MME mean change and associated **(b)** signal-to-noise ratio; **(c)** Proportion of variance per factor for the MME mean change: GCM (yellow), GHM (green), Residual (red).

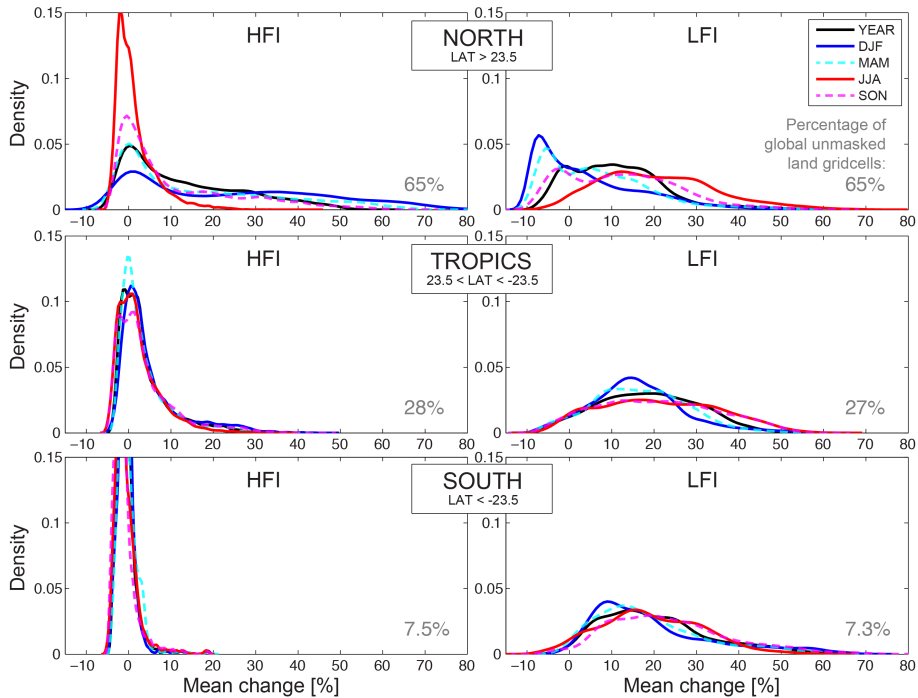


Figure 2. PDFs of mean changes in high (HFI) and low (LFI) flows, annually and per season (DJF and JJA) for North, Tropics, and South latitude bands.

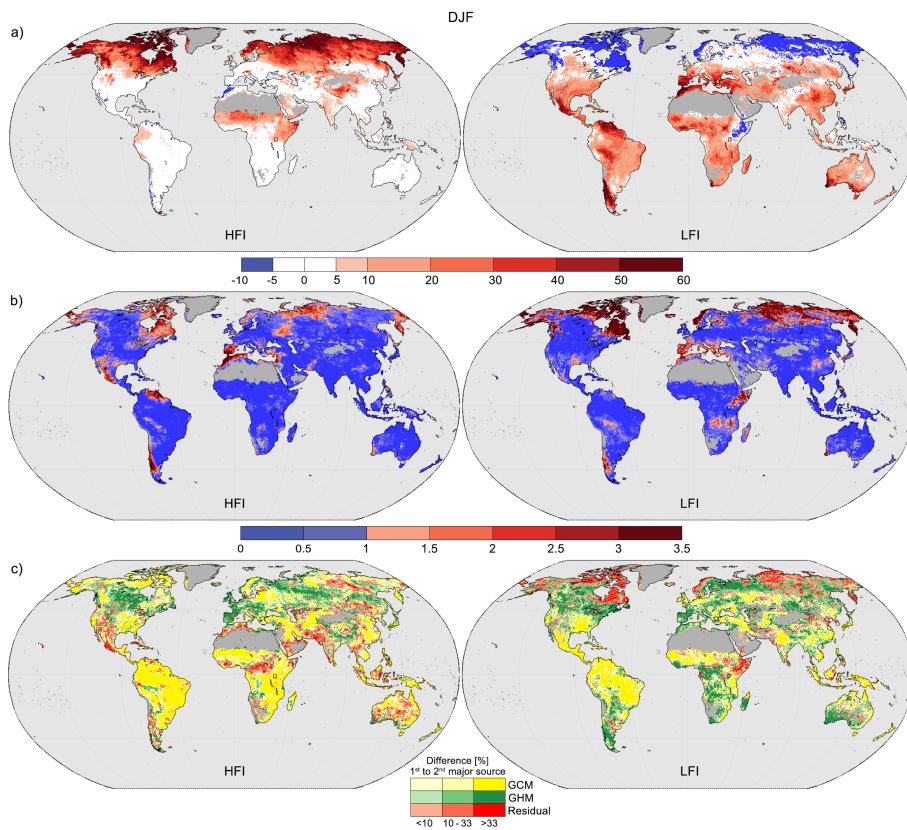


Figure 3. Same as Fig. 1 for the season DJF.

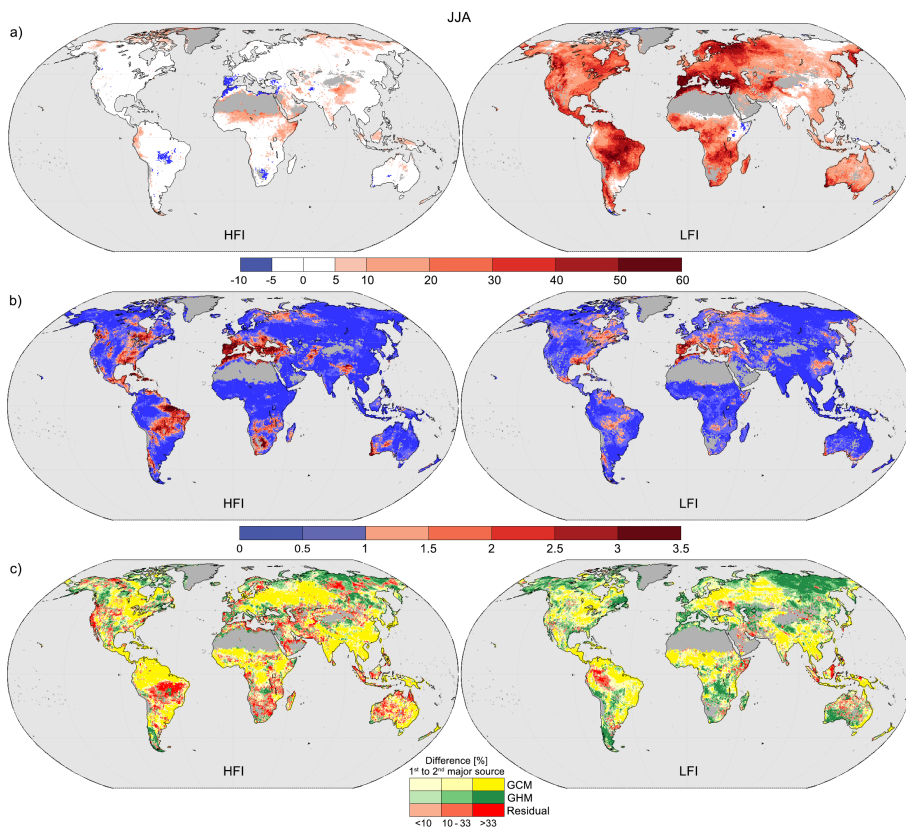


Figure 4. Same as Figs. 1 and 3 for the season JJA.

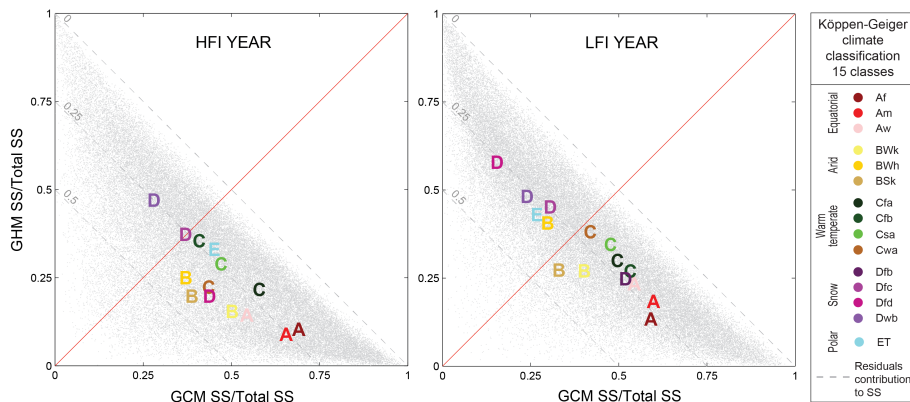


Figure 5. ANOVA sum of squares (SS) of the 2 factors (GHM y axis; GCM x axis) divided by the total sum of squares (TSS) for all grid cells as gray dots; and for each Köppen–Geiger climate region (15 most represented), as region letters shown at the medians of the region’s GCM SS/TSS as x coord and of the regions’s GHM SS/TSS as y coord.

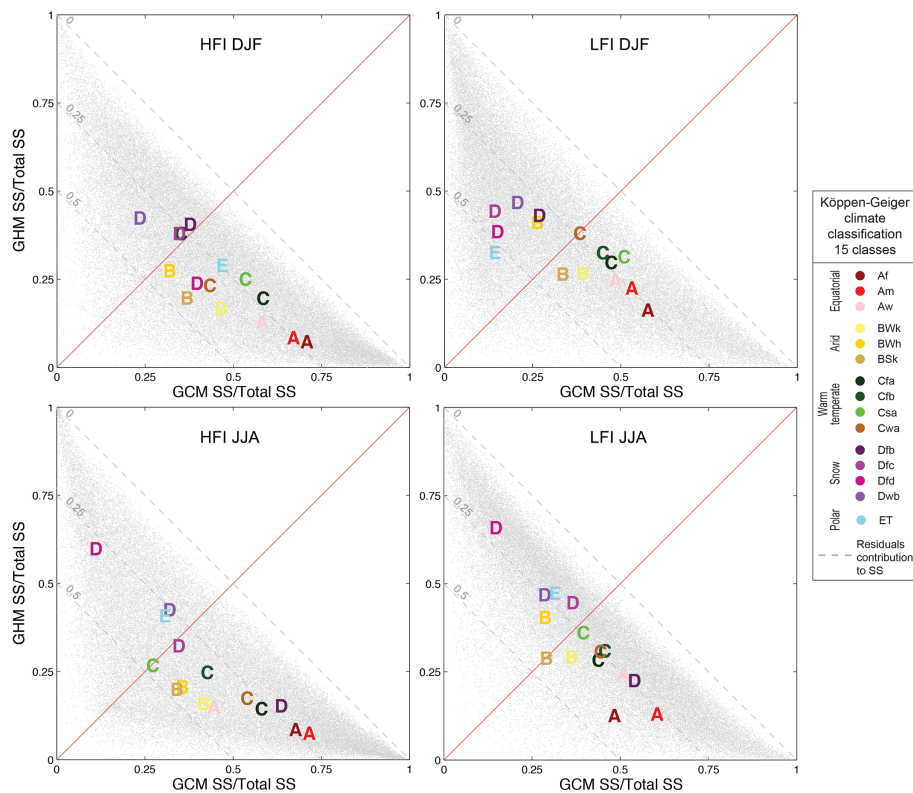


Figure 6. Same as Fig. 5 for the seasons DJF (top) and JJA (bottom).

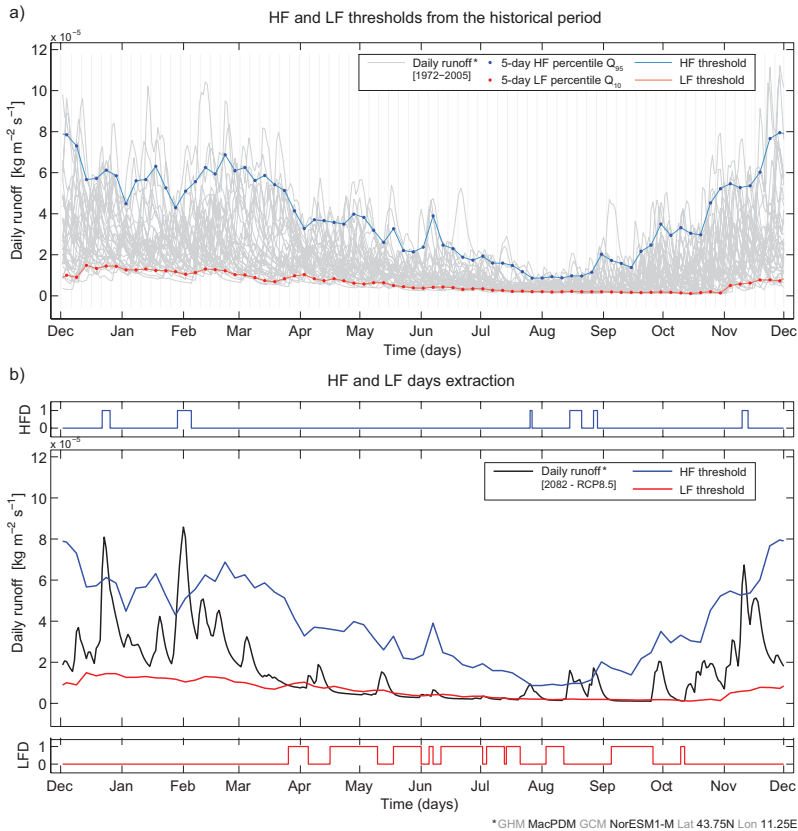


Figure B1. Schematic of HFD and LFD extraction (days under high and low flows): **(a)** Daily varying threshold curves for HF and LF from 5 day percentiles calculated over the historical period; **(b)** High and low flows days extraction for a given year. As an example, runs of a Southern European gridcell (Lat 43.75° N, Lon 11.25° E) from **(a)** historical (December 1971 to December 2005) and **(b)** RCP8.5 (2082) periods of the MacPDM/NorESM1-M were used for this figure.

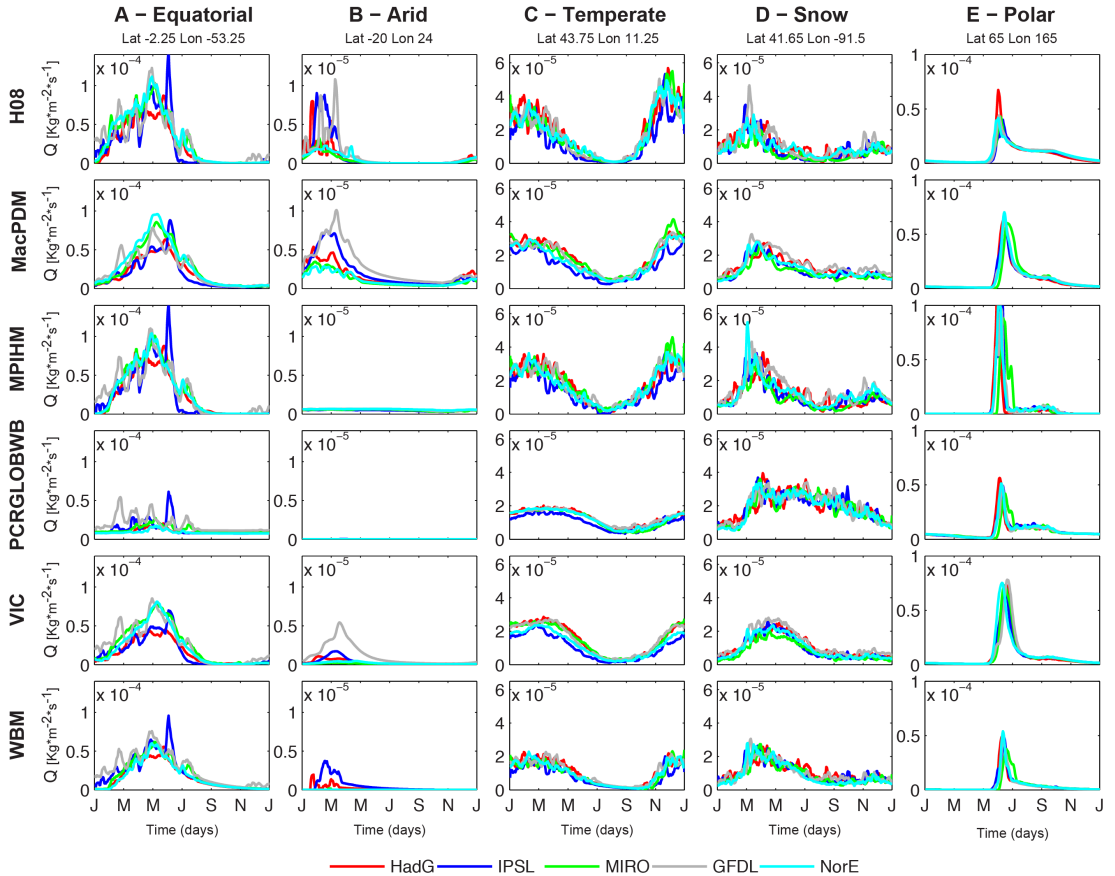


Figure B2. Inter-annual dynamics in mean daily runoff (smoothed with a 7-d moving average) relative to the period 1972-2005 for selected Köppen-Geiger region gridcells: A-Tropical (-2.25° N, -53.25° E) Northern Brazil; B-Arid (-20° N, 25° E) Botswana; C-Temperate (43.75° N, 11.25° E) Central Italy; D-Snow (41.65° N, -91.5° E) Central U.S.; E-Polar (65° N, 165° E) N-Eastern Russia.

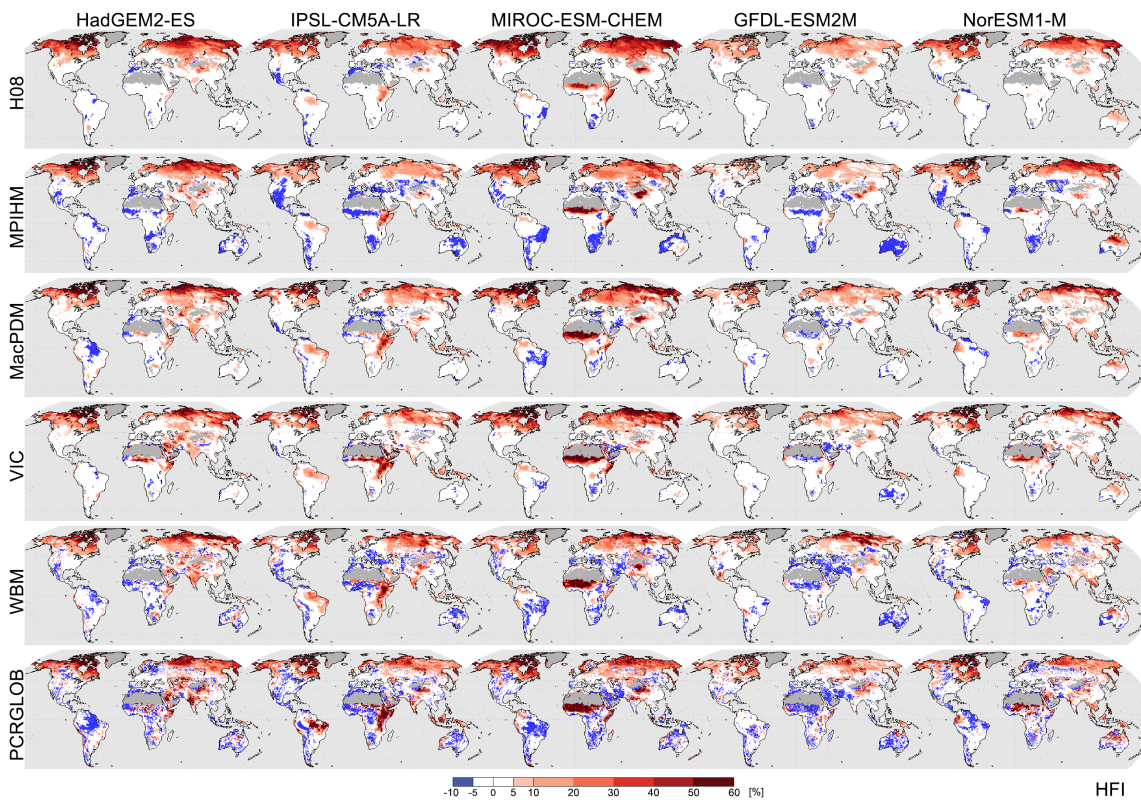


Figure B3. Same as Fig. 1a for individual GHM (row) – GCM (column) combination, for HFI.

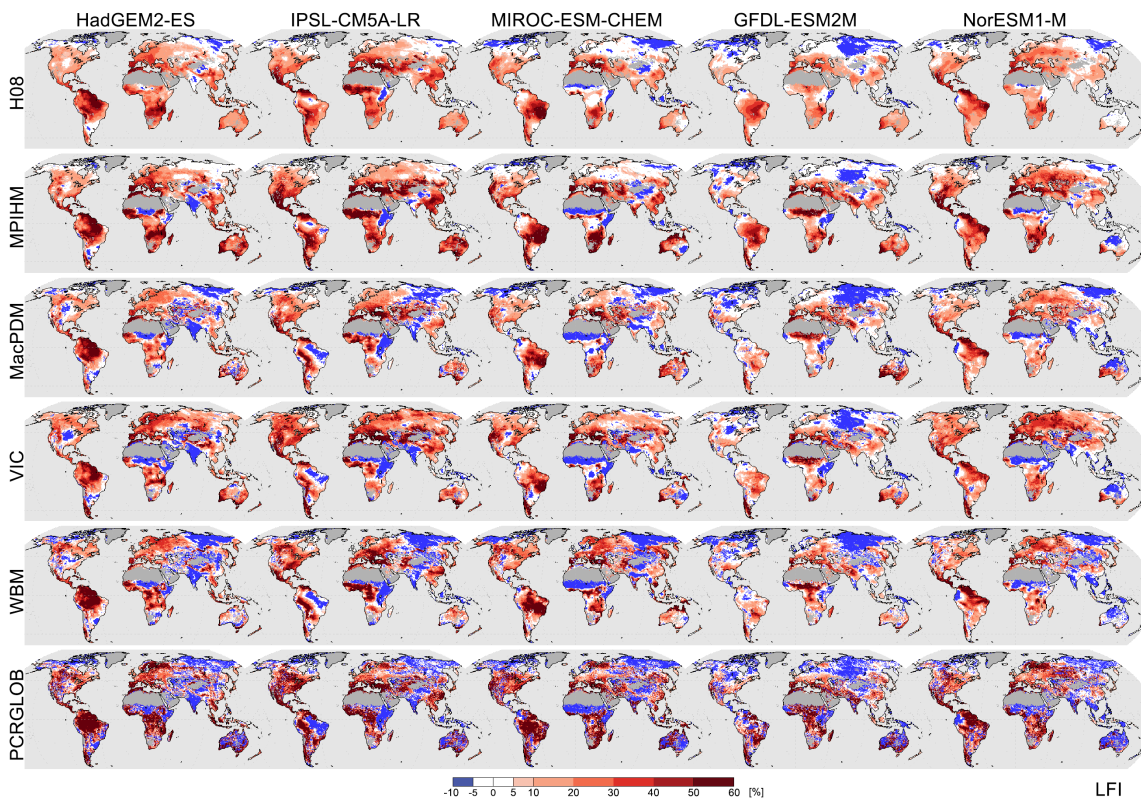


Figure B4. Same as Fig. 1a for individual GHM (row) – GCM (column) combination, for LFI.