Relies to the comments of Anonymous Referee #2

Authors' replies are in blue color and revised sentences are in italics.

General Comments:

The work presented in the manuscript with the title 'A Global Approach to Defining Flood Seasons', aims to develop a methodology that allows defining spatial and temporal characteristics of major flood seasons globally with the help of daily stream flow simulations.

As the work reuses of already existing and published modelled global streamflow data, the central scientific contribution of this manuscript is the development of an approach that allows defining flood seasons globally. With this in mind, it would be valuable for the authors of the manuscript to focus more on how they define floods seasons and to compare their results with other already existing flood season indicators.

We thank the anonymous reviewer for the positive comments and further critical comments that we believe have enhanced the overall quality of the manuscript.

Besides focusing on the major flood season globally, the study does also briefly consider minor floods seasons. As the authors point out, the study of minor flood seasons has not obtained much attention at a global scale and therefore merits further investigation.

Therefore, the manuscript would benefit from extending the scope to minor flood seasons not only at the local scale (i.e. with an example from East Africa as presented in the manuscript) but also to the global scale, which would also better match the overall global scope of the manuscript.

Thanks for the compliments on minor flood seasons. We have extended the scope of minor flood seasons to global scale. Please see below responses below on minor flood seasons.

In general, the paper need to be clearer about the different meaning and usage of the terms 'peak month (PM)', 'flood season (FS)' (is it always the 3 months (i.e. PM +-1 month)?), and 'PAMF' (Percentage of Annual Maximum Flow (AMF)). Sometimes these variables are used almost interchangeable. On this matter, also see the specific the comments below.

As the Reviewer surmised, the flood season (FS) is always the 3 month period with the peak month (PM) as the center month. We have addressed and clarified this in the specific comments below.

With the general comments above and the specific comments mentioned below, I recommend thoroughly revising the manuscript, as there are several instances that require further clarification, discussion, corrections, and amendments from the authors.

Overall, the paper is well written and has the potential to be of interest to the readership of HESS. Therefore, I suggest resubmitting the manuscript after a major revision.

Again, the authors thank the Reviewer for their constructive review.

Specific Comments:

Section Abstract:

P 4596 L4-6: The authors argue in their abstract that 'forecasting systems in the order of months to seasons are a rarity' and that 'dominant flood seasons must be adequately defined' for prediction and disaster preparedness.

I agree that there is a shortage of long-term forecasting systems; however, I would say that in general the flood regime and therefore the flood prone seasons of rivers are locally (the scale at which preparation for disasters take place) well know.

In addition, I presume that the hydrological model, from which the discharge data has been obtained, performs very different at different scale. This is of particular importance, as ungauged basins, for which this type of information would be useful, are often smaller than the grid scale of the model.

Therefore, it is suggested changing the reasoning/focus of the abstract (and also the introduction and conclusions), as the approach to define flood seasons (of 3 months in length) that has been developed here has only a marginal connection with disaster preparedness and flood forecasting.

We agree with the reviewer's comment. The original motivation of this research was to understand temporal variability of global streamflow in order to improve global-scale flood prediction. However, as suggested, the major findings of this article are not directly connected to disaster preparedness and flood forecasting, thus we will revise/restructure the abstract, introduction and conclusion to focus more on flood seasons and their hydrological applications.

Section 1 Introduction:

The introduction focuses on long-range seasonal forecasts for guiding decision-making, seasonal predictability of streamflow impacts and the need for linking atmospheric indices with streamflow predictions at global scale. From this introduction, I would expect a paper that aims to PREDICT streamflow patterns, which is very different to the actual scope of the paper.

For that reason, I think the introduction should focus more on the actual topic (i.e. a data based approach concerned with the identification of flood seasons and the second objective of extending the approach to already existing globally modelled streamflow).

We agree with reviewer's comments. We have revised the introduction to highlight the identification of flood seasons and tempered the discussion of prediction (which will be the subject of future work.)

To put the work presented in the manuscript into the context of already existing studies of flood seasonality and global streamflow characteristics, the authors may find the following articles useful: For previous work on different method of identifying/classifying flood seasonality see for example Ouarda et.al. 2006, Liu eta 2010 or Chen et.al. 2013. For more information on how the manuscript fits in the context of or differs to other global studies of streamflow characteristics see for example Dettinger and Diaz 2000 or Beck et.al. 2015.

Thanks for the valuable references on flood seasons and global streamflow characteristics. We have read and added these as appropriate to the manuscript.

Section 2 Data description:

2.1 Streamflow stations

P 4599 L 5-7: The current selection of the dataset cannot really be considered 'global' and has a particular bias towards to certain regions (particularly northern hemisphere).

Please provide further explanation on the how the stations were selected (see also comments below).

Thanks for the comments on this issue. To clarify, the hydrologic (gridded) model does have full spatial coverage, however the station data is, as the reviewer suggested, less well dispersed. Additionally, not all stations have equal record lengths or record quality, however for verification aspects, comparing model outputs with observations is obviously critical. We have provided specific procedures for selecting stations below,

Does 'having at least 20 years of continuous daily streamflow data' mean that all stations that had one measurement missing were excluded, or was a threshold on missing data applied?

Please further explain how the selection criterion 'continuous daily streamflow data' influenced the spatial coverage of the data.

We apologize for the misunderstanding. We have selected stations with more than 20 complete years not necessarily continuous 20 years. If missing values are found, corresponding years are excluded. The criteria of 20 complete years and locations on model's river network have, admittedly, reduced the number of stations, particularly for the southern hemisphere, and specifically the Africa and South America, because of relatively poor observed data.

To what period does the 'at least 20 years' refer to? Same the stream flow simulations (1958-2000)? Please specify.

Yes, the same period with model's simulation (1958-2000) was used. We have clarified this in the manuscript.

With a less stringent selection criterion ('having at least 20 years of continuous daily streamflow data'), could one have obtained a better compromise of spatial coverage and data quality?

The Reviewer makes a good point. Yes, if the station data criteria is relaxed, it is possible to add more stations and have better spatial coverage, however there are not many stations that were "close" to meeting the criteria, then dropped. Stations that did not meet the criteria were typically quite short in record length or had many missing data points. So in the end, the authors believe that revising this strict selection process is unlikely to make a significant difference in the number of stations added.

Please add a paragraph further elaborates on these choices, as this step is crucial in determining the amount of data and spatial coverage available for method validation.

We have changed P.4599, line 5-7 to:

Daily streamflow observations utilized in this study are from the Global Runoff Data Centre (GRDC, 2007.) For comparing flood seasons between simulation and observation, stations *located along with the model's drainage network are considered. Station records that are missing even short periods may effect how a flood season is defined, thus we have excluded years with any daily missing values from stations. In this study, a minimum of 20 hydrological years required for a station to be retained. Globally, 691 stations from all continents except Antarctica were selected with data more than 20 complete years, with upstream area ranging from 9,539 to 4,680,000 km² and period of record between 20 and 43 years from 1958 to 2000 (Figure 1.) Although this criteria is admittedly quite strict (no missing daily data), relaxing the criteria does not add a significant number of stations.*

2.2. PCR-GLOBWB

P 4599 L20: The authors mention that the model was forced with input data from ERA-40, which 'were subjected to a number of corrections'.

Please specify how these corrections might or might not influence the model output.

The WATCH project performed these corrections to reproduce more realistic atmospheric data. Thus, it may affect the model's performance in simulating more accurate streamflow. For specifying these corrections, we have changed P.4599, line 18-21 to:

The WATCH forcing data were originally derived from the ERA-40 reanalysis product (Uppala et al., 2005), and were subjected to a number of corrections including elevation correction, time-scale adjustments of daily values to reflect monthly observations, and corrections for varying atmospheric aerosol-loading and separate precipitation gauge corrections. It is possible that this may have some minor effect on streamflow simulation, likely providing more realistic outcomes. Full details of corrections are described in Weedon et al. (2011).

Please also discuss/analyse the influence of the hydrological model and the grid cell size on the ability of the model to generate the magnitude of hydrological extremes, which will be used as a key variable for the definition of the flood season using the volumetric-based threshold approach.

The authors agree that proper realization of hydrologic extremes in the model is important and validation of these characteristics is necessary. To support this, we have provided references illustrating the model's ability (as this is work performed prior to our analysis.). We have provided the following sentence after P.4599, line 12:

The PCR-GLOBWB model has not been calibrated, thus simulation results may be biased and uncertain at course spatial resolution, however it has the ability to provide long time-series of streamflow globally, which has is sufficient to estimate long-term flow characteristics with spatial consistence (Winsemius et al., 2013). Additionally, this model has been validated in previous studies in terms of streamflow (Van Beek et al., 2011), terrestrial water storage (Wada et al., 2011) and extreme discharges (Ward et al., 2013), indicating model performance.

3. Defining flood seasons

3.1 Methodology for defining grid-cell scale flood seasons

P 4061 L1-7: It is pointed out that it is important to consider not just the magnitude but also volume to define a flood season and that the authors therefore adopt a volume-based threshold technique. However, the authors then select the 'streamflow exceeding the top 5 % of the FDC', which is related to magnitude. If this is the case, it is not clear why the need for/use of a volume-based measurements is highlighted here and several times throughout the document.

We apologize for the misunderstanding that the volume-based threshold method is used to record streamflow occurrences (volume) based on the top 5% of the FDC (magnitude). For clarification, we have changed P.4601, line 1-7 to:

Therefore, to define the FS, and specifically the PM, both volume and magnitude aspects need to be considered (Javelle et al. 2003). To do this, we adopt a volume-based threshold technique. This technique is similar to a streamflow volume-based method in terms of capturing the Julian day by which a fixed percentage of the annual streamflow volume has occurred (Burn, 2008), however it also applies this fixed percentage across the entire streamflow record and records points where streamflow volume surpasses it, drawing from the prescribed threshold concept in the POT method. Here we select streamflow surpassing the top 5% of the flow duration curve (FDC) across all years (1958-2000) as the threshold for considering a high streamflow level, as commonly adopted in threshold approaches (Burn, 2008; Mishra et al., 2011.)

Additionally, please elaborate on the decision process of selecting the '5% threshold', as on the previous page the importance on selecting the 'proper threshold for POT' highlighted. Have other thresholds been tested and what was the outcome?

We selected the top 5% of the flow duration curve as a threshold as this is commonly used to consider high streamflow level in threshold approaches (p.4601, line 3-7). We did not test other threshold levels, but rather compared to other indices (section 3.2) to show that the volume-based threshold method is the best method to consider magnitude and volume simultaneously. That said, considering varying threshold levels could be an interesting aspect of future work.

P 4061 L10-12: From the description, it appears that after identifying the peak month, the flood season is defined as the month before and after the peak month. Is this the case or is the flood season related to the three month with the highest number of days above the 5% streamflow threshold?

We apologize for the misunderstanding. The FS is defined as the month before and after the PM. The PM is the key outcome here, however for the future (intended) prediction work, a seasonal approach will be undertaken, thus we have also defined the FS. For clarification, we have changed P.4600, line 2-5 to:

To identify spatial and temporal patterns of dominant streamflow uniformly, we design a fixed time window for representing flood seasons globally. Here we define major flood seasons as the 3-month period most likely to contain dominant streamflow and the annual maximum flow. The central month is referred to as the Peak Month (PM) and the full 3-month period is referred to as the Flood Season (FS.) Specifically, we define PM first, and then define FS as the period also containing the month before and after the PM. This approach is performed for both observed (station) and simulated (model) streamflow to gauge performance.

I could imagine a situation similar to the synthetic streamflow data used in Figure 2, with August (105 days) being the peak months but June with 60 days (instead of the 25 days used in the example)) and July with 75 days. Resulting in the peak month being off centre.

It needs to be clarified, if such a situation had been considered and if not how that will influence the results (including the calculation of the index 'Percentage of Annual Maximum Flow (PAMF)').

The intent of the Percentage of Annual Maximum Flood (PAMF) metric created for this analysis directly addresses the reviewer's concern. After defining the FS, it is evaluated in terms of how many of the annual maximums are contained within that 3-month period (see next response). So it certainly is possible that the PM could be "off center", however, generally the PAMF value will be highest in the PM. A good example is Figure 11(a) (please see updated Figure 11(a) below). Here, PM (April) could be regarded as being off center from largest streamflow volume (AMJ), however, PAMF value in April is higher than May. (Here, monthly PAMF is the PAMF value calculated at each month, please see below reformulated Eq. (1)).

P 4601 L 20-24: The index PAMF has been created to 'evaluate' the identified flood seasons. Therefore, it is suggested interpreting the 'high' or 'low' values of PAMF in that regard, (e.g. a high PAMF indicates a well represented Flood Season (FS), a low PAMF indicates poorly identified FS?). Additionally, as the index has been created to evaluate the defined FS, please also give an indication of what is considered by the authors of being a good or less acceptable value (i.e. what percentages are considered good ? And for the discussion of the results for what regions is the approach used to define flood seasons not working).

We appreciate the reviewer's comments on this. For clarifying this, P. 4601, line 13-14 has been changed to:

To evaluate the defined FS objectively in terms of how many of the annual maximum flows are contained, we develop a simple evaluating statistic called the Percentage of Annual Maximum Flow (PAMF).

To address what may be an acceptable value for PAMF, we have classified FS PAMF ranges. We have changed P. 4601, line 20-24 to:

For example, a high PAMF indicates that the FS is highly likely to contain the annual maximum flood each year. In contrast, a low PAMF indicates that the timing of the annual maximum flow is more likely to vary temporally, and may be a result of bimodal seasonality, consistently high or low streamflow throughout the year, streamflow regulated by infrastructure or natural variation. In this study, we subjectively classify FS PAMF values as: high = 80-100% PAMF, low = 60-80%, and poor = 40-60%.

P 4601: Generally, after highlighting the advantages of the POT approach and the disadvantages of the annual maximum flows (AMF) (such as) it is not clear to me why the PAMF method uses AMF to evaluate the defined flood seasons. Has other values instead of the AMF been considered, and if so why has the AMF been chosen?

Yes, we used a threshold approach only for defining PM (volume-based threshold technique) for considering streamflow volume and magnitude. We had considered using threshold-based index for evaluating FS which may show higher scores than the AMF-based index, however we selected the AM-based statistic because it is more objective to evaluate how many of the annual maximum flows are contained in FS, compared to subjective threshold-based statistic. Additionally, this AM-based statistic is more applicable to further analysis of minor FS.

3.2 Classification techniques

P 4602 L 12-14: Please further elaborate why ‘1-7 days favour identifying flood magnitude, while 15- 30- days favour identifying flood volume’.

Thanks for the comments on this issue. As the reviewer’s comments, it is be a subjective classification. Here we classified a QAM and Q7 index for flood magnitude and Q15 and Q30 for flood volume by comparison with full length of PM (30 days). For clarifying this, we have changed P. 4602, line 12-14 to:

Compared to the full length of PM (30 days), the flow-based classification techniques with a shorter time component (1-7 days) favor identifying flood magnitude while the techniques with longer time components (15-30 days) favor identifying flood volume.

P 4602 L 16-21: I am having difficulties in understanding what this section means. Can you please rephrase and explain for what reason ‘they may be considered slightly superior’.

The PAMF value is developed for evaluating FS in terms of how many of the annual maximum flows are contained. Therefore, if classification techniques define a FS differently at the same station, the technique showing the highest PAMF value should be superior to others in terms of containing annual maximum flows. For clarification, we have changed P. 4602, line 17-21 to:

The PAMF is also useful for comparing classification techniques’ performances when they define PM differently at the same location. This occurs at 45% and 40% of stations for observation and simulation, respectively. The classification technique having the highest PAMF most often for those stations may be considered slightly superior in terms of containing more annual maximum flows in their defined FSs. The volume-based threshold technique has the highest PAMF values by at least 2% of stations more than other techniques for simulation, and at least 1% of stations more than other techniques without Q_AM for observation.

How can one calculate the PAMF for the other classification techniques?

The PAMF is calculated in the same way using Eq. (1). So, for example, if all classification methods define the same PM, all PAMF values are the same. However, if PMs are defined differently, the PAMF values would be unique to their corresponding PMs.

Can equation (1) be reformulated to be more generally applicable (see also comment on monthly PAMF below).

We have provided it in a general format. P.4601 line15-18 has been changed to:

$$PAMF(i) = \frac{\sum_{j=i-1}^{i+1} nAMF(j)}{\sum_{k=1}^{12} nAMF(k)},\ 1 \le i \le 12 \qquad (1)$$

where $nAMF(i)$ denotes number of annual maximum flows that occurs in i month during entire records. In Eq. (1), when i is 1 (Jan), $i-1$ in the summation is 12 (Dec), and when i is 12 (Dec), $i+1$ is 1 (Jan). Here the PAMF provides the percent of time the annual maximum flows occurs in the defined FS across the evaluation period.

General Comment on Section 3.1 and 3.2:

It is not clear to me, why these two sections are separate. I would expect to evaluate the PM that has been identified with the 5% threshold approach together with the other classification techniques and then pick the best indicator (i.e. here apparently the PM) for further analysis. In addition, if applicable compare the performance with other seasonal indicators that have been publish in other studies before and explain why the approach here is superior to the other methods (otherwise the new approach would not be needed).

Having first the 'Methodology for defining grid-cell scale flood seasons' and then having a separate section on '3.2 Classification techniques' is confusing.

Therefore, I suggest combining these two section together with an in depth analyses on the flood season classification approaches (e.g. how do global maps of differ?).

We agree with reviewer's comments. Section 3.2 has now been merged into section 3.1.

3.3 Methodology for defining sub-basin scale flood seasons

P 4603 L 21-22: I understand that under certain circumstances, the PAMF can be useful to indentify managed dams or reservoirs, but if the dams are managed in accordance with what is considered the 'natural flow regime', this will not help. Additionally, why not use the dataset mentioned a few lines above to find the location of the dams? Without the factual knowledge of the presence/absence of a dam, one will have difficulties in determining if the low values of the PAMF indicator obtained from the modelled data are due to management or due to difficulties of the model to represent the hydrological characteristics of that region.

We chose not to use the reservoir dataset explicitly, because if a station's seasonality is not affected by an upstream reservoir (as suggested by the author), it might contribute in defining the sub-basin PM, and we would want to retain that information. We have checked reservoir locations against downstream stations for many locations to verify our assumption. There may be cases of low PAMF due to presence of a dam that we have not discovered, but a cursory evaluation did not reveal this to be a point of concern. For clarifying this, we have changed P.4603, line 21-23 to:

The PAMF, as previously defined, can aid in identifying stations affected by upstream reservoir by showing low PAMF values. This is applied with the assumption that reservoir flood control disperses the annual maximum flows across months rather concentrated within a few months (e.g. akin to natural flow.)

4. Verification of selected flood seasons

P 4604: I suggest adding the characteristics of the data obtained from the DFO (such as available period...) here, instead of having it in section 4.2 (P 4606), where I would focus on discussing the results.

We agree with the reviewer's suggestions and have provided specific information about DFO records here. P.4604, line.24-25 has been changed to:

First, the model-based PMs are verified by comparing with observation-based PMs at station and sub-basin scales. Also, historic flood records from the Dartmouth Flood Observatory (DFO) are used to compare globally defined PMs to actual flooded areas spatially and temporally.

Specifically, we used the following information from DFO: start time, end time, duration and geographically estimated area at 3,486 flood records during 1985-2008.

4.1 Observed vs. Modelled flood seasons

P 4605 L 6: How are the temporal differences calculate? Is it based on the (central?) peak months or on the entire 3 month long flood season? I.e. if I have an observed Flood season June to August, and a modelled season September to November, is the difference three or just one months?

Yes, the difference is between PMs, so in the example provided by the reviewer, it would be 3 months difference. For clarification, we have changed P. 4605, line 4-6 to:

For comparing modeled PMs to observations, the defined PMs and calculated PAMF are represented globally at the station scale (Figure 4-5) and sub-basin scale (Figure 6) with temporal differences of PMs (modeled PM – observed PM).

I would suggest to calculate the differences not for the FS but for the PM (if the PM is centred in 3 month flood season (see also discussion on the definition of FS above)) and the PAMF respectively and then add a panel showing the differences in Figure 4-6 respectively directly, allowing a direct comparison (instead of having them separately in Figure 7).

We agree with the reviewer that this will be easier to interpret and have changed the name to temporal difference of PM, and provided direct comparison maps (temporal difference maps). Please see updated Figure 4 and 6 below.

P 4605 L 6: In Figure 7 (P 4624), the temporal differences are shown.

However, the colour scale of the Figure seems to omit basins with differences larger than +- 4 months (see catchments highlighted with red boarders in the Figure excerpt below)!

Please check again, why these catchments are not shown. If these catchments actually have such extreme differences in the FS, please do not omit them from the discussion in section 4.1.

This is an important part of the analysis, which is currently not apparent to the reader and should be highlighted and discussed!

Thank you for pointing out this. We have corrected this issue. Please see responses on corresponding figures below.

P 4605 L 7: Is there is a mix-up with the % of stations used in the text to describe the bar plot. 62% and 44% seem to refer to the % of stations of the entire dataset but I rather think that the percent should read according to the height of the bars ~ 35% and ~50%. Please check.

Thanks for correcting this. We have checked all values mentioned in this section. P.4605, line 7-9 has been changed to:

For example, in the United States and Canada, 38% of stations and 51% of sub-basins produce identical PMs, growing to 82% of stations and 93% of sub-basins when considering a ±1 month temporal difference (Figure 7.)

P 4606 L 6-7: I'm not sure how the authors come to the conclusion of 'Europe exemplifying a constant-flow region'. From my knowledge of the flood hydrology in Europe, I would say that most of the regions in Europe have a well-defined seasonal flood regime. Could the authors please better explain how this had been concluded.

We made this conclusion based on the streamflow simulation around northeastern Europe, however there were some biased simulations (please see biased PM (large difference in PM) in updated figure 6(c) or 8 below). After considering the reviewer's comments, it is difficult to generalize about the entire Europe, thus we have removed this sentence.

General Comment on Section 4.1:

As the main aim of the paper is to define flood seasons globally, I recommend a more in depth analysis of the obtained differences in the PM or the FS.

For example, it would be valuable to analyse if the differences between the observed and modelled PM and FS are systematically linked to station/sub-basin characteristics such as catchment size, latitude/ longitude or altitude. The results will then give a better fell on the reliability of the modelled PM and FS not only in light of possible human influences (e.g. dams or reservoirs) as discussed in the manuscript

Thanks for the comments. We had planned to do more analyses discovering characteristics on difference of PM between simulation and observation, however, we thought any other conclusions about them would be beyond the scope of section 4.1 which is to show the performance of modeled PM compared to observation. Additionally, we have focused more on the minor flood seasons that might physically explain significance of flood season. We do agree that this is relevant and interesting, however, and will consider further detailed analysis in this direction in the future.

4.2. Modelled flood seasons vs. actual flood records

As mentioned before, I would move the characterisation of the DFO data into section 4 and focus here on a more quantitative assessment of the differences.

P 4606 L 23: To me the there is no 'striking similarity' between the DFO and the modelled data. Maybe if the authors summarise the gridded model data to the same sub-basin scale as the DFO, similarity may becomes more apparent. I therefore suggest to also providing some sort quantification (not only qualitatively discussing the maps), before calling it 'striking'.

It is possible to perform a quantitative assessments for comparing DFO records and cell-based PM using GIS applications, however, the DFO data used in this study was generated by different sources, and polygons in the DFO map are not spatially observed areas, but spatially *estimated* areas. Also, each polygon has qualitative information, such as main cause, mortality, damage and so on, therefore we thought any spatially drawn conclusions based on a quantitative comparison between globally defined PM and DFO records could be unreasonable or unjustifiable. We indeed explored this initially.

The previous color code was not continuous from Dec to Jan, thus we have changed it to be continuous (we appreciate reviewer's comments.) The updated PM and DFO maps (please see updated Figure 8 and 10 below) are similar, and mostly dissimilar areas have low PAMF values, indicating "unstable" Annual Maximum Floods. In section 5, we have defined the minor FS that helps to explain low PAMF values at the corresponding regions, and compared global major and minor FS maps and monthly DFO maps (please see updated figure 13 below).

> 5. Defining minor flood seasons
>
> Defining minor flood seasons is a very relevant research topic that has obtained little attention in global studies, as the authors point out (P 4607 L 14-21).
>
> This is where I would see a great contribution of this manuscript in advancing the scientific understanding of flood seasons.
>
> Unfortunately, this aspect is only covered briefly and appears to be appended to the main analysis, currently with limited added value.

The authors appreciate the reviewer's comment and thus we have extended the defined Minor FS to global-scale, and provided more analyses.

> P 4607 L 23: Please explain how monthly PAMF values were calculated. (I suggest using a more general formula for equation 1).

We have provided general formula for Eq. (1). Here the monthly PAMF values are calculated by Eq. (1) at each month. For clarifying this, we have changed P. 4607, line 22-24 to:

To detect noteworthy minor flood seasons globally, we classify streamflow regimes by climatology and monthly PAMF value, which is calculated using Eq. (1) at each month (Figure 11.)

> P 4608 L 14-16: after describing in length the methods used to define the minor PM, the authors only show an example of East Africa. Here, I would have expected a global map showing regions where such minor flood seasons are existing and if possible indicating the PM as well on global maps.

We have provided more analyses on minor flood seasons and extended the scope to global-scale as suggested. P.4607, line 23 – P.4608, line 16 have been changed to (please also see new figure 12 below):

To detect noteworthy minor flood seasons globally, we classify streamflow regimes by climatology and monthly PAMF value, which is calculated using Eq. (1) at each month (Figure 11.) Classifications include unimodal, bimodal, constant, and low-flow. The unimodal streamflow climatology has high values of PAMF around the PM; the bi-modal classification is represented by two peaks of PAMF; both constant and low-flow classifications represent low values of PAMF between months. Distinguishing between bi-modal and other classifications is nontrivial. For example, upon initial inspection of the constant streamflow classification (both climatology and monthly PAMF, Figure 11 (c)), it could be mistaken for a non-dominant bi-modal distribution. In other words, bi-modal streamflow could be detected correctly or incorrectly, depending on how to define bi-modal streamflow. We adopt the following criteria to differentiate bi-modal streamflow from uni-modal, constant and low-flow conditions.

- *The low-flow classification is defined for annual average streamflow less than 1 cms.*
- *The major and minor PMs must be separated by at least two months in order to prevent an overlap of each FS (3-month.)*
- *If there is a peak in monthly PAMF values around major FS, it is regarded as potential minor PM.*
- *If the sum of both major and minor PM's PAMF is greater than 60% (minimum of 29 out of 43 annual maximums fall in one of the FS), it is defined as bi-modal streamflow.*

For considering potential flooding, the minor PM is identified by the secondary peak of monthly PAMF rather than the magnitude or shape of streamflow. Also, the minor FS is not defined when a major PM's PAMF is greater than 80% (minimum of 35 out of 43 annual maximums), indicating a robust uni-modal streamflow character (Figure 11 (a)). The sum of both major and minor PM's PAMF is used to determine the significance of both FSs in terms of containing annual maximum flows; a high value of the joint PAMFs (80-100%) indicates that both FSs are significant (Figure 11 (b)), moderate values (60-80%) imply less significance with some probability of being classified as constant streamflow (Figure 11 (c)); low values (50-60%) are likely constant or low streamflow (Figure 11 (d)). After defining the major PM, the minor PM is identified globally with the corresponding joint PAMF values (Figure 12), and the minor FS is also defined as the month before and after the minor PM. In Figure 12, minor FSs are evident in the tropics and sub-tropics and spatially consistent with bi-modal rainfall regimes discovered by Wang (1994). Well-known bi-modal flood seasons are also defined in East Africa (second rainy season in winter) and Canada (rainfall-dominated runoff in autumn) with high joint PAMF values (80-100%) indicating strong significance of both FSs. Minor FSs are also associated with bi-modal rainy seasons, for example the major FS (NDJ) and minor FS (MAM) in Central Africa consistent with the latitudinal movement of the ITCZ, intra-Americas' major FS (ASON) due to the major rainy season and minor FS (AMJJ) due to minor rainy season (Chen and Taylor, 2002), and coastal regions of British Columbia in Canada and southern Alaska's minor FS (SOND) due to wintertime migration of the Aleutian low from the central north Pacific (Figure 12). Also, distinct runoff process from different climate systems can induce a bi-modal peak within a large-scale basin, such as the upstream sections of the Yenisey and Lena river systems in Russia where their major FS (AMJ) is dominated by snowmelt and thawing and minor FS (JAS) is spurred on by the Asian monsoon period. The same mechanism produces minor FSs around the extents of the Asian summer monsoon with high significances (90-100% of sum of PAMFs) (Figure 9 and 12). Moderate minor FSs include, for example, the southern United States (Texas and Oklahoma) bi-modal rainfall pattern (AMJ and SON) and southwestern United States (Arizona) where summer major FS (JJA) is produced by the North American monsoon and winter minor FS (DJF) is affected by the regional large-scale low pressure system (Woodhouse, 1997). Southeastern Brazil's summer major FS (NDJF) and post-summer minor FS (AMJ) are dominated by formation and migration of the South Atlantic Convergence Zone (Herdies, 2002; Lima and Satyamurty, 2010). In central and eastern Europe, the major FS (FMAM) and minor FS (JJA) are defined as moderate (60-80% of joint PAMF values for central Europe and 70%-90% for eastern Europe); for northeastern Europe the major FS (MAM) and minor FS (NDJ) are have high joint PAMF values (80%-100%.).

For the major FS and minor FS with joint PAMF values exceeding 60% (Figure 13), flood records (DFO) occurring over more than one month are counted in each month based on the reported duration of all records. Although one distinct flood event dominate a monthly DFO record, strong similarity is evident between the FSs and monthly flood records (Figure 13.) The *minor FSs with high PAMF values corresponding well with the observed DFO flood records are in East Africa (notable bi-modal streamflow), intra-Americas and Northern Asia; only a few reported flood records occur in the minor FSs in high latitudes. The minor FSs with moderate PAMF values are evident in the southern US, southeastern Brazil and central Europe.*

6. Conclusions and discussions

P 4608 L 20-23: The authors highlight that the streamflow model was evaluated 'to define dominant and minor flood seasons globally'. This has only been partly archived for the case of the dominate season, not for the minor seasons (see comments for section 5 above).

We have now provided global-scale major and minor flood season maps.

P 4609: As already mentioned in the comments to the introduction, the conclusion of the manuscript focuses on many other aspects surrounding the topic of prediction and links to global and regional climate links, which has little to do with the main focus of the manuscript in the current form.

Therefore, I would suggest, revising this section.

We agree with reviewer's comments. We have revised it in the manuscript.

Comments to Tables and Figures

P 4616: Table 1 Please add the '5% ' to the threshold column heading.

Thanks, we have added it.

P 4618: Figure 1. When printing the manuscript on my printer (printer-friendly version form the HESSD website), the background polygons are not visible. Please check.

Thanks, we have changed the background color to be darker for all figures.

P 4621: Figure 4: When printing the manuscript, the colour code for the points does not allow me to identify the different months properly. For example, I cannot distinguish points indicating April from March or May. Please use a different colour scheme.

Thanks, we changed the color scheme for Peak Month (Figure 4, 6 and 8) and DFO months (Figure 10). Please see updated figures below.

P 4624 Figure 7: Adjustment of the plotting procedure is necessary to accommodate basins that have differences larger than 4 months and therefore currently are hidden and not visible at all.

P 4624 and P 4624: Please add to the Figure captions, what the meaning or + and - are. (i.e. Do positive red values mean that the observed PM/FS is x months earlier and negative blue values indicate that the observed PM occur x months later?)

Thanks for catching this mistake. The difference of PM is calculated by PM (simulation) – PM (observation). We have provided the inadvertently omitted stations and sub-basins, and also captions. Please see updated Figure 4 (c) and 6 (c) below.

References.

Ouarda, T., Cunderlik, J., St-Hilaire, A., Barbet, M., Bruneau, P. and Bobée, B.: Data-based comparison of seasonality-based regional flood frequency methods, Journal of Hydrology, 330(1), 329–339, 2006.

Liu, P., Guo, S., Xiong, L. and Chen, L.: Flood season segmentation based on the probability change- point analysis technique, Hydrological Sciences Journal, 55(4), 540–554, 2010.

Chen, L., Singh, V. P., Guo, S., Fang, B. and Liu, P.: A new method for identification of flood seasons using directional statistics, Hydrological Sciences Journal, 58(1), 28–40, 2013.

Dettinger, M. D. and Diaz, H. F.: Global characteristics of stream flow seasonality and variability, Journal of Hydrometeorology, 1(4), 289–310, 2000.

Beck, H. E., de Roo, A. and von Dijk, A.: Global maps of streamflow characteristics based on observations from several thousand catchments, Journal of Hydrometeorology, published online, doi: 10.1175/JHM-D-14-0155.1; 2015.

Updated figures

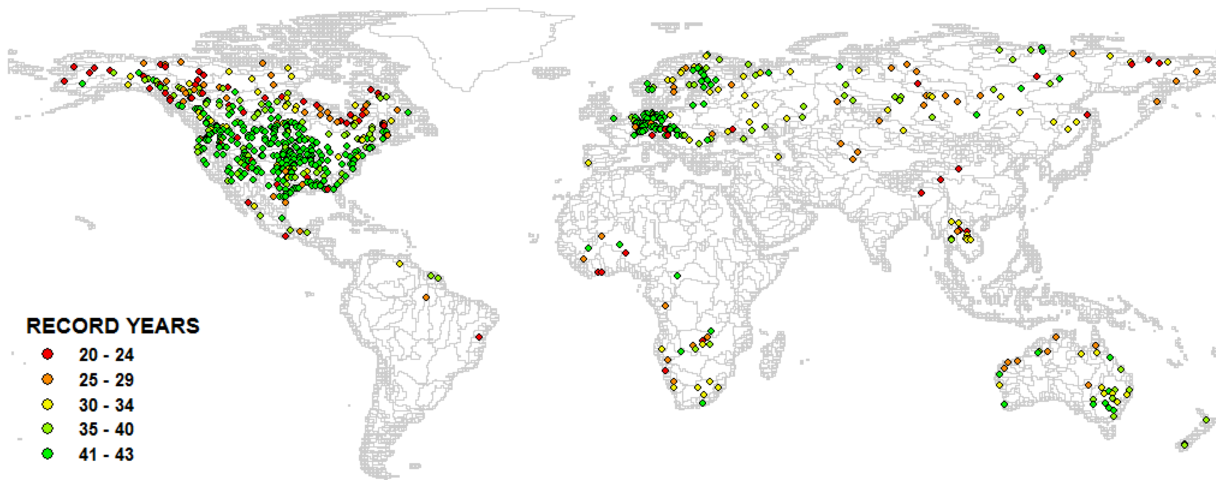

Figure 1. Location of 691 selected GRDC stations with corresponding number of years per station. Background polygons are world sub-basins based on 30′ drainage direction maps (Döll and Lehner, 2002) with separation of large basins (Ward et al., 2014).

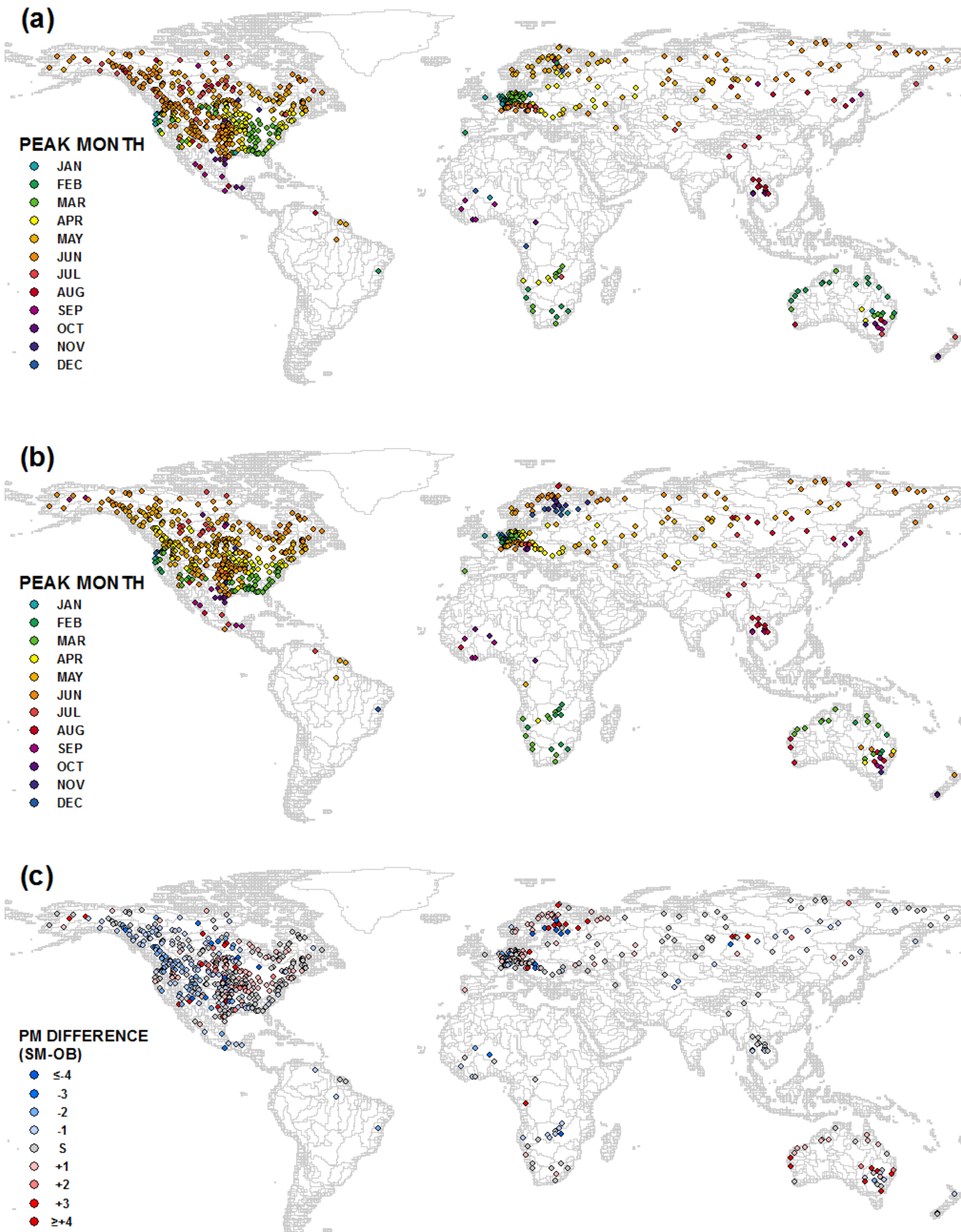

Figure 4. Peak Month (PM) for flooding as defined by (a) 691 GRDC observation stations, (b) simulated streamflow at associated locations and (c) Temporal difference (SM-OB, number of months) in PM between observations and model outputs.

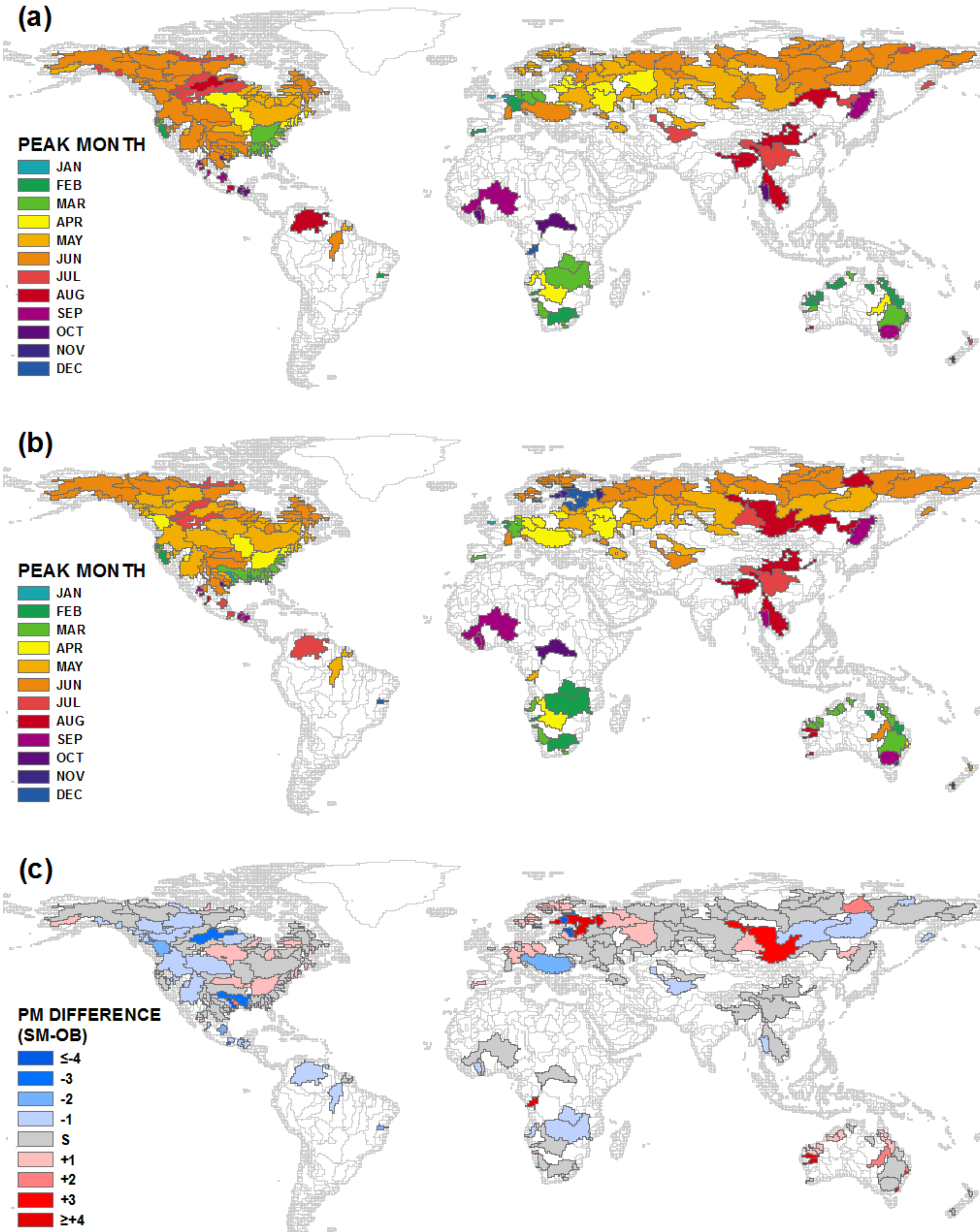

Figure 6. Peak Month (PM) for flooding by sub-basin as defined by (a) 691 GRDC observation stations, (b) simulated streamflow at associated sub-basins and (c) Temporal difference (SM-OB, number of months) in PM between observations and model outputs.

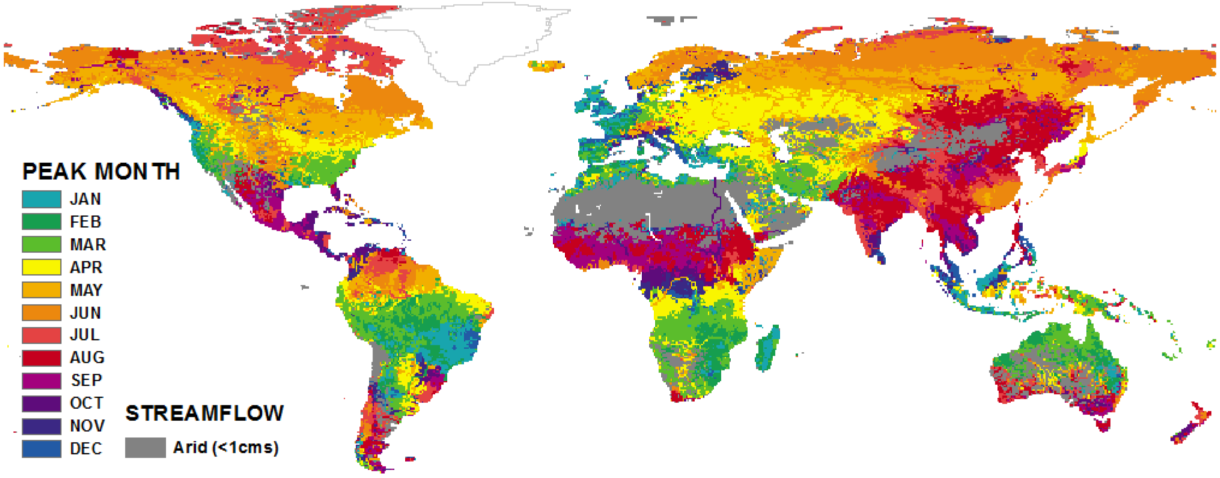

Figure 8. Peak Month (PM) for flooding as defined at all modeled grid cells.

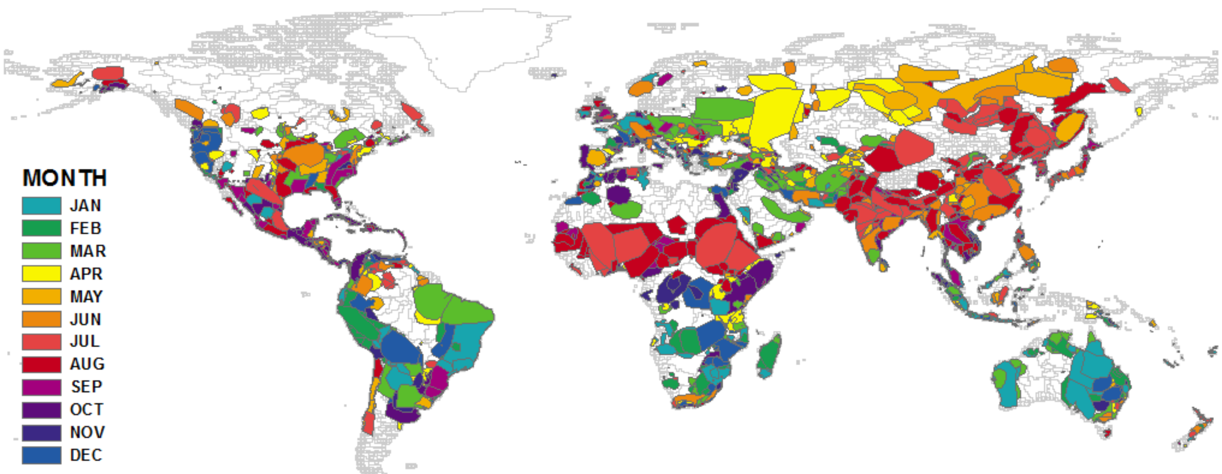

Figure 10. Archive of major flood events globally from the Dartmouth Flood Observatory (DFO) over 1985-2008.

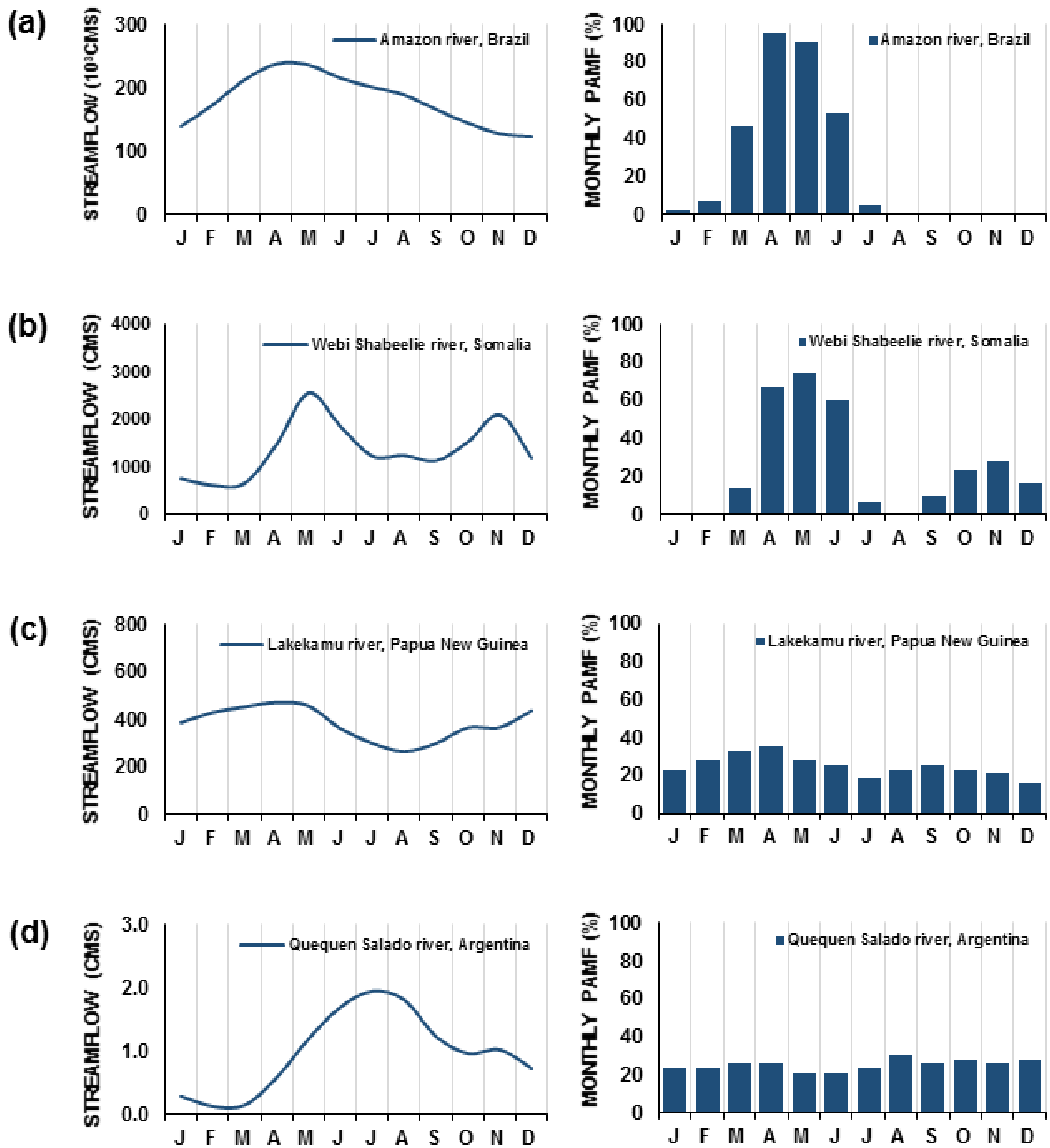

Figure 11. Model-based streamflow climatology (left) and corresponding monthly PAMF (right.) Types and locations are: a) uni-modal streamflow – At Bom Lugar, Amazon river, Brazil, b) bimodal streamflow – At Saacow, Webi Shabeelie river, Somalia, c) constant streamflow – At Terapo Mission, Lakekamu river, Papua New Guinea and d) low-flow – At La Sortija, Quequen Salado river, Argentina.

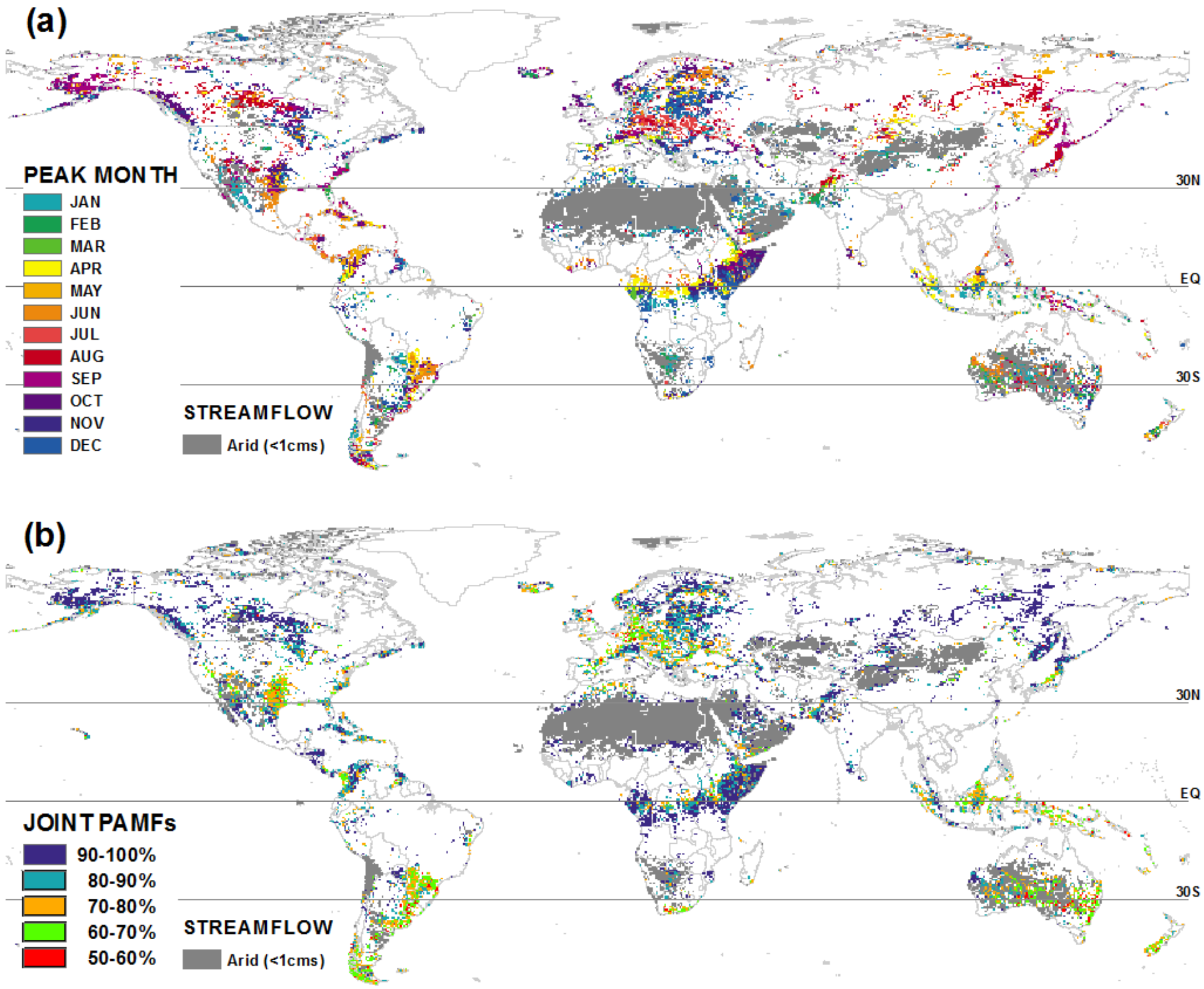

Figure 12. (a) Minor Peak Month (PM) for flooding as defined at detected grid cells and (b) joint PAMFs of major and minor PMs at corresponding cells.

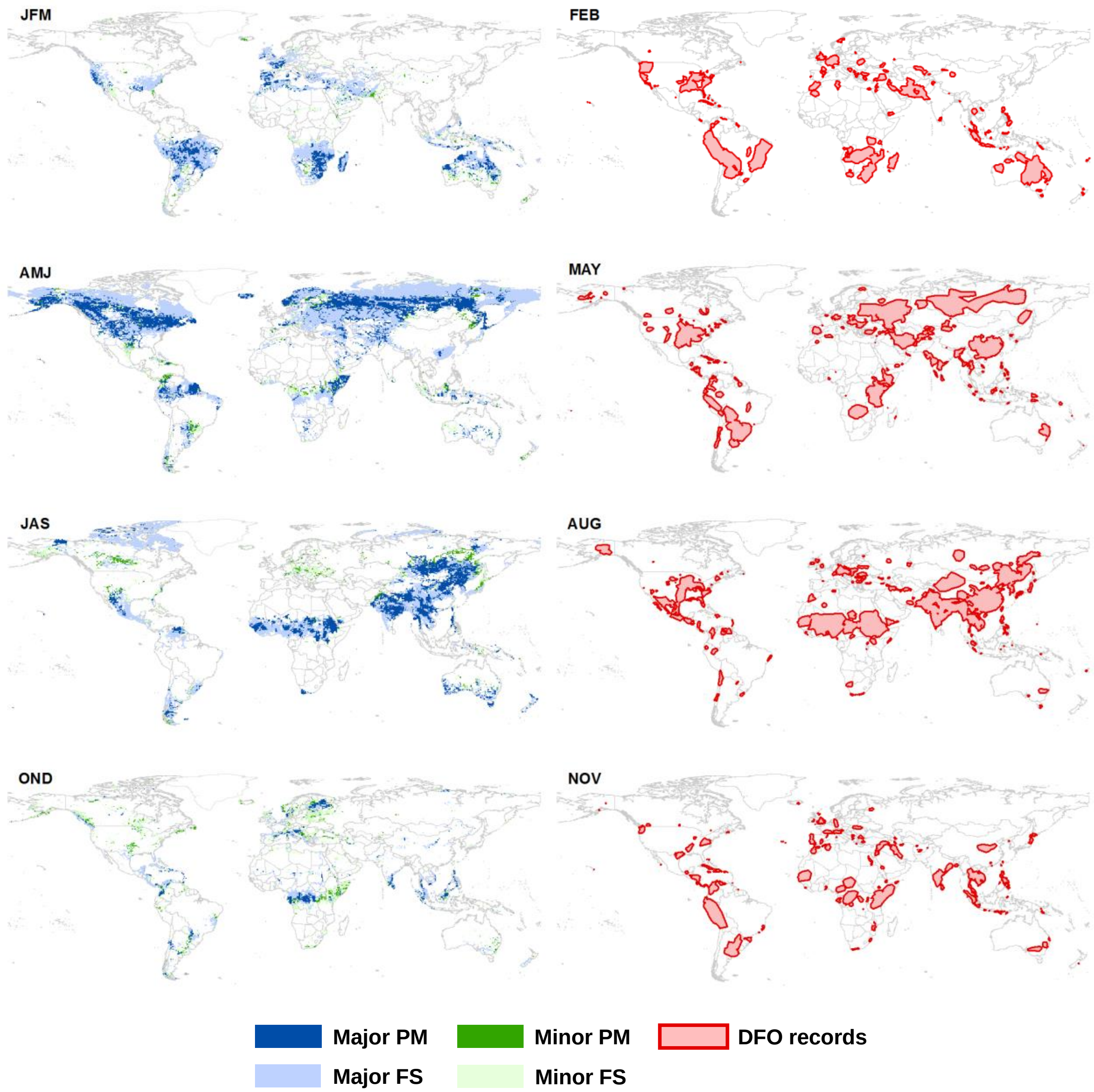

Figure 13. Total flood seasons (left); peak month of major and minor FSs (dense color) and post-month of prior FS and pre-month of next FS (light color.) Monthly accumulated actual flood records (DFO) during 1958-2008 (right.)