Manuscript No.: hess-2015-140
Original Title: A Global Approach to Defining Flood Seasons
Authors: Donghoon Lee, Philip Ward and Paul Block

Dear Editor and Reviewers,

We would like to thank the Editor and three Referees for again carefully reviewing our manuscript and the additional critical and helpful comments raised by Referee #2. We have revised the manuscript accordingly and detailed these changes in the response, below. Referee comments are presented in indented text, our responses are in blue, and revised sentences are in italics. Page and line numbers are given with respect to the last revised manuscript (submitted 03 Aug 2015). The revised manuscript includes underlined (added text) and crossed out (deleted text) for easy reference.

Based on Referee #2's suggestions, the following major changes have been made:

(1) The title has been changed to "Defining High-flow Seasons using Temporal Streamflow Patterns from a Global Model."

(2) The new method proposed here has been compared with other common methods at various threshold levels.

(2) We have revised the results and discussion sections to be more balanced by adding shortcomings of this approach.

We believe the suggestions and ensuing revisions have clearly resulted in a higher quality manuscript and eagerly look forward to any further comments.

Yours sincerely,

Donghoon Lee

Relies to the comments of Anonymous Referee #2

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

General Comments:

The manuscript has been revised thoroughly by the authors and some of the previous comments of the reviewers. This improved the clarity of what the authors had actually been doing, but also raised some more points to comment.

Overall, there are still some key aspects that need to be addressed before the paper can be published (together with some other minor aspects mentioned below).

1. The title of the paper has been changed to 'Defining Flood Seasons Globally using Temporal Streamflow Patterns'.

However, with the clarifications and explanations on the methodology provided by the authors after the first review, this title does not seem appropriate!

The methodology presented enables the identification of the 'high flow' seasons and NOT the 'flood season'! So the using 'flood seasons' in the title is misleading.

For many regions in the world, the 'real floods' do actually not occur during the high flow season (in the manuscript called 'flood season.')

This is actually also corroborated in several of their results. E.g. Figure 9, in which (although difficult to tell from the colour code used in that plot) more than 1/3 to 1/2 of all pixels fall in the evaluation class of 'low' to 'poor' (based on their own 'subjective' classification scheme (P7 L 17-18)).

Therefore, the paper is far from 'defining flood seasons globally

Additionally, I would suggest, that the title should indicate that the 'global' scale of the paper comes from model output, as the actual data available for checking the results has strong spatial biases.

We agree with the referee's comment. The title has therefore been changed to "Defining High Flow Seasons using Temporal Streamflow Patterns from a Global Model." This eliminates confusion between high flow and flood season, as they may not be synonymous as suggested by the referee. We have also added the words "global model" to clarify the model scale. Finally, the term "Flood Season (FS)" in the manuscript has been changed to "High-flow Season (HS)" to match the new title.

2. The authors highlight their new technique to define 'flood seasons'. However, no formal testing of the new technique is performed, in which the technique would be compared to other already well established techniques that aim to define the flood season.

Generally, the method is similar to a 'peak over threshold approach' but instead of considering independent peaks, all daily flows above a volume based threshold are used to define the 'flood season' (which is actually the 'high flow season', or the 'high flow spell').

> Therefore, the method simply identifies the month with the highest number of days above a flow quantile (here the upper 95% percentile). It is not clear why, the authors call it a new 'volume and magnitude' based technique. Additionally, a proper testing of the sensitivity of the results obtained to the threshold selected is essential when presenting a 'new method'.

The referee makes a couple good points that we have addressed. First, the focus of the manuscript is not to present a "new method", but rather through our process we identified a potentially superior approach which is really simply an extension of existing methods. Thus we have tempered any text that highlights this as a new method. Secondly, we initially omitted any details on the performance of this method in comparison to common methods, and now have added those results in the revised manuscript.

The Volume-Based Threshold (VBT) method presented here is really just an extension of the common volume-based method. Both methods record the days (dates) that the threshold (percentile) is exceeded, and the total volume for the days exceeded. Thus duration aspects are also captured. The VBT is an extension, or alternative, in that it applies the threshold over the entire time-series (available record) concurrently instead of on a year-by-year basis. In other words, for the 95^{th} percentile, instead of annually calculating the 95^{th} percentile, it is calculated using the entire period of record. The common volume-based method thus records events every year surpassing the threshold, however for the VBT approach, this need not be the case. This approach emphasizes capturing the key peaks (as in a peak over threshold approach.) For clarification, we have changed P.6 L 22-26 to:

This technique is similar to a streamflow volume-based technique in terms of capturing the days (Julian dates) when streamflow exceeds the pre-defined threshold (percentile of flows) and associated volume (Burn, 2008.) The major difference, however, is that the VBT applies the threshold over the entire time-series (available record) concurrently instead of on a year-by-year basis. In other words, for the 95th percentile, instead of annually calculating the 95th percentile, it is calculated using the entire period of record. The common volume-based technique thus records events every year surpassing the threshold, however for the VBT approach, every year need not have a peak above the threshold. This approach emphasizes capturing the key peaks across the entire available time-series (as in a peak over threshold approach.) VBT thus contains both volume and timing characteristics for defining the Peak Month (PM.)

The performance of the VBT method is tested against both the common volume-based approach and the POT approach. Four volume-based methods with 1%, 3%, 5% and 10% of annual volume (V01%, V03%, V05% and V10%, respectively) and three POT methods averaging 1, 2 and 3 peaks per year (POT1, POT2 and POT3, respectively) are compared. Further, a sensitivity analysis of VBT method is added evaluating and comparing four flow thresholds (percentiles of annual flows at 1%, 3%, 5% and 10%.) We have updated P.7 L 20 – P.8 L17 to:

The VBT technique is compared with the common volume-based technique and POT technique to gauge performance. Four volume-based durations, namely V01%, V03%, V05% and V10% and three POT techniques averaging 1, 2, and 3 peaks per year (POT1, POT2 and POT3, respectively) are selected. For the V01% technique, the HS is simply centered on the PM containing the largest number of occurrences of the top 1% of annual streamflow volume across the total years available. The V03%, V05% and V10% techniques are similar to the V01%

approach, respectively using 3%, 5% and 10% of annual streamflow volume. Comparatively, techniques with a shorter time component (1-3% of annual volume) favor identifying the PM by peak timing while techniques with longer time components (5-10% of annual volume) favor identifying the PM based on duration and peak volume. The VBT technique is an attempt to bridge these two criteria. For the POT techniques, independence criteria is applied to avoid counting multiple peaks from the same event (Institute of Hydrology, 1999.) For example, two peaks must be separated by at least three-times the average rising time to peak, and minimum flow between two peaks must be less than two-thirds of the higher one of the two peaks. More details of independence criteria are described in Lang et al. (1999.)

An analysis examining sensitivity of selected threshold levels for the VBT technique is also undertaken. Performance of thresholds representing 1%, 3%, 5% and 10% exceedance across the entire period of record, named VBT1%, VBT3%, VBT5% and VBT10%, respectively, are compared.

To compare techniques and thresholds, the PMs are defined at the 691 selected stations and associated model grids. The locations where the PMs differ (by at least one technique) are of most interest. This occurs at 61% of stations and 54% of associated grids. Cross-correlations of PM between the four common volume-based techniques clearly indicate the tendency of the defined PM to shift from peak timing dominated to peak volume dominated as the time component increases (Table 1.) Correlation between VBT techniques and volume-based techniques are quite similar and consistent (0.82-0.86 and 0.84-0.86 for observed and simulated streamflow, using VBT5%; Table 1), preliminarily indicating some success in capturing both timing and volume properties, while correlation between the VBT techniques and POT are less strong (0.78-0.81 and 0.79-0.83 for observed and simulated streamflow, respectively, using VBT5%; Table 1.) The PAMF is also useful for comparing techniques, such that the technique having the highest average PAMF typically contains more annual maximum flow events in their defined HSs. The VBT5% is superior to other VBT and POT techniques for both observed and modeled streamflow, having the highest PAMF values, however the volume-based techniques indicate similar or even slightly better performance than VBT5% (Table 2.) This is not unexpected as the volume-based techniques are designed to capture annual peak flows on a year-by-year basis, whereas the VBT records significant peaks across the full time-series, and may “miss” annual peaks in some years in which that peak is small relative to all peaks throughout the available record. Thus VBT tends to select PMs that contain the most significant peaks overall, and subsequently have the highest potential for capturing probable flood seasons for flood-prone basins, a desirable outcome for this study. To illustrate this in the context of the PAMF, if all years are ranked for each location based on the annual peak flow, and the top 50% (half) are retained, the PAMF actually favors the VBT approach, surpassing the volume-based approach by 5-6% for PMs and 2-3% for HSs.

Finally, techniques may be evaluated by comparing the temporal difference (number of months) between model-based and observed PMs; closer is clearly superior. The VBT3% and VBT5% techniques produce the greatest degree of similarity between model-based and observed PMs (81% of stations having ±1 month difference; Table 3.) Overall, the VBT technique demonstrates superior performance as compared with the POT techniques by all comparisons. The VBT technique is also on par or slightly superior to the common volume-based technique, especially considering the 5% threshold; thus, the remainder of the analysis is carried out utilizing the VBT5% technique only.

Table 1. Cross-correlations of Peak Month (PM) for each classification technique for observed and simulated streamflow where stations having different PMs.

Classification Technique		VBT1%	VBT3%	VBT5%	VBT10%	V01%	V03%	V05%	V10%	POT1	POT2	POT3
Observed	VBT1%	1.00										
	VBT3%	0.90	1.00									
	VBT5%	0.85	0.94	1.00								
	VBT10%	0.79	0.86	0.91	1.00							
	V01%	0.82	0.82	0.82	0.81	1.00						
	V03%	0.81	0.84	0.83	0.84	0.89	1.00					
	V05%	0.81	0.85	0.86	0.85	0.86	0.92	1.00				
	V10%	0.80	0.84	0.85	0.87	0.83	0.88	0.96	1.00			
	POT1	0.78	0.78	0.78	0.74	0.76	0.77	0.76	0.74	1.00		
	POT2	0.74	0.78	0.78	0.78	0.80	0.80	0.82	0.81	0.81	1.00	
	POT3	0.77	0.81	0.81	0.80	0.80	0.81	0.83	0.81	0.86	0.93	1.00
Simulated	VBT1%	1.00										
	VBT3%	0.87	1.00									
	VBT5%	0.83	0.95	1.00								
	VBT10%	0.80	0.88	0.90	1.00							
	V01%	0.86	0.85	0.84	0.84	1.00						
	V03%	0.87	0.86	0.85	0.83	0.92	1.00					
	V05%	0.87	0.88	0.85	0.84	0.90	0.97	1.00				
	V10%	0.82	0.87	0.86	0.85	0.83	0.89	0.92	1.00			
	POT1	0.80	0.83	0.83	0.81	0.83	0.86	0.86	0.82	1.00		
	POT2	0.78	0.81	0.80	0.79	0.79	0.83	0.83	0.82	0.92	1.00	
	POT3	0.80	0.81	0.79	0.80	0.80	0.83	0.84	0.81	0.92	0.95	1.00

Table 2. Average PAMF of each classification technique for modeled and observed where stations having different PMs.

Section	VBT1%	VBT3%	VBT5%	VBT10%	V01%	V03%	V05%	V10%	POT1	POT2	POT3
Observed	60.8%	61.7%	62.0%	62.0%	63.4%	63.6%	63.0%	62.5%	60.8%	59.1%	60.6%
Simulated	63.5%	64.5%	64.7%	63.5%	65.1%	64.8%	64.9%	64.1%	63.1%	60.3%	61.9%

Table 3. Percentage of stations according to the difference in PMs between modeled and observed streamflow at each classification technique.

Difference in PMs	VBT1%	VBT3%	VBT5%	VBT10%	V01%	V03%	V05%	V10%	POT1	POT2	POT3
Same	39%	39%	40%	42%	38%	39%	40%	42%	38%	36%	38%
≤±1 month	80%	81%	81%	80%	78%	79%	79%	79%	75%	75%	77%
≤±2 month	90%	91%	91%	90%	89%	90%	89%	89%	87%	87%	88%
≤±3 month	94%	95%	95%	95%	94%	95%	95%	95%	93%	93%	94%

3. Additionally, the editor's requests 'What are the advantages to defining the new measure PM (and FS), in relation to existing published measures of flood seasonality? What shortcomings did the previous existing published measures have, and to what extent do your new measures overcome those shortcomings? have not been addressed!

Indeed we did not properly address this comment and apologize for missing it the first time around. The main shortcoming of published literature identifying global-scale flood seasons has to do with peak timing; it is predominantly a bi-product of analyses that focus on flood season identification from a streamflow amplitude perspective, and as a result, is somewhat subjective. Granted, defining objective peak timing is not the focus of most analyses, however interpreting peak timing can be problematic due to varying seasonal patterns (e.g. bi-modal distribution, constant or low flow areas, etc.) These studies also often cluster regions into large homogeneous areas as defined by streamflow amplitude – and peak timing along with it – however defining "average" peak timing at such a large scale may not be justified. In lieu of clustering or aggregation, here we identify high-flow seasons (PM and HS) by capturing annual peak timing using the VBT method at the cell and sub-basin scale, presenting an approach focused on streamflow temporal patterns rather than pattern of amplitude. The new measure of PM (and HS) coupled with the model grid scale provides much higher resolution peak timings globally than previously presented (often at large basin scale or subcontinental scale.) The performance measure introduced here (PAMF) is also a new contribution relating the models ability to capture high flow season timing. These advantages are also helpful for identifying less-dominant but important seasons (minor high-flow seasons) that possess similar characteristics to the high flow season (e.g. bi-modal annual cycle), another unique contribution of this work. This leads to better temporal characterization and understanding of flood potential, causation, and management, particularly in ungauged or limited-gauged basins. For clarification, P.4 L 3-16 has been changed to:

In general, these studies define high streamflow or flood seasons subjectively based on the relationship between dominant streamflow amplitude patterns and large-scale climate drivers/patterns, and delineate large-scale homogeneous regions correspondingly. Defining high flow season timing is essentially a bi-product of these analyses, and may be problematic due to varying seasonal patterns (e.g. bi-modal distribution, constant or low flow areas, etc.) not captured at the large-scale delineation. There is also typically no distinguishment between minor and high flow seasons. In some cases, these minor seasons (e.g. resulting from bi-modal precipitation distribution) can produce high flow or flood conditions, and are thus of interest to identify. Here we identify high-flow seasons by capturing annual peak timing using a volumetric technique at the cell and sub-basin scale, presenting an approach focused on streamflow temporal patterns rather than pattern of amplitude. The new measure of PM (and HS) coupled with the model grid scale provides much higher resolution peak timings globally than previously presented (often at large basin scale or subcontinental scale.) The performance measure introduced here (PAMF) is also a new contribution relating the models ability to capture high flow season timing. These advantages are also helpful for identifying less-dominant but important seasons (minor high-flow seasons) that possess similar characteristics to the high flow season (e.g. bi-modal annual cycle), another unique contribution of this work. This leads to better temporal characterization and understanding of flood potential, causation, and management, particularly in ungauged or limited-gauged basins.

The authors only compare their method to other approaches, by cross-correlating the identified peak months with other techniques.

Although the correlations applied are similar (Table 1) (which would be expected by using the same dataset and techniques that aim to identify similar features of the flow regime), a correlation between the different classification techniques cannot be used to justify the superior performance of their new method!

Correlation should never be confused with causation!

Instead, differences in the obtained correlation with any of the other classification techniques could actually mean that these techniques are superior in their performance in capturing the peaks!

The referee makes a good point – and the authors agree. The purpose of using correlation is not to demonstrate the superiority of the VBT method, but rather to illustrate the similarity between the VBT method and all common volume-based methods (V01%, V03%, V05% and V10%) By doing this, we intend to further demonstrating that the VBT method captures aspects of both short and long time components (peak timing and peak volume properties). For clarification, we have changed the corresponding text appropriately. Please see the revised text in query 2, above.

I urge the authors to follow the well-established research approach of first testing all these techniques and then select the most appropriate technique for the rest of the analyses (which might be the new method, but maybe the older techniques perform better (one cannot tell from the current manuscript)). Instead of coming up with a new method first and then not thoroughly evaluating if the new method actually is better!

We fully agree with referee's comments, and were remiss in failing to include them initially. We now compare the VBT method with other traditional approaches, including the common volume-based method at four durations and the POT method at three thresholds. Additionally, the VBT method is subjected to a sensitivity analysis at four threshold levels. Overall the proposed VBT5% method tends to show slight superiority. Please see the revised text in query 2, above.

Specific Comments:

Section Abstract:

P2 L3-4: Please rephrase, as the sentence currently gives the idea that only the new approach of defining flood seasons is 'objective' and not the other methods. I think the current approach is as 'objective' as the other methods. So I would restrain from using the word 'objective' here.

The authors agree. The word 'objective' has been removed from the abstract and other relevant places in the document.

P2 L9: I disagree that the defined flood seasons represent well the actual flood records from DFO. This is only achieved when the minor secondary flood seasons are included later in the manuscript! Please rephrase.

The authors agree and have changed P.2 L 8-11 to:

Minor high-flow seasons are also defined for bi-modal flow regimes. Identified major and minor high-flow seasons together are found to well represent actual flood records from the Dartmouth Flood Observatory, further substantiating the model's ability to reproduce the appropriate high-flow season.

> P2 L12-15: This is a false claim. The identified seasons (which are the peak month +- 1 month) do certainly not help to improve the understanding of flood frequency, trends and interannual variability. Please remove.

The authors agree. The sentence has been removed.

> P4 L3-12: In this paragraph, the shortcomings of the previous studies are presented. High emphasis is placed on the issue of clustering. However, for most of the studies this is not the main aim, instead they also show very distinct seasonality patterns. Addition, the other studies are criticised for 'not being representative of local scale conditions ' and that the current study is addressing 'basin and even grid cells'. However, the analysis of this study is not 'local' either and the approach used to define a sub-basin's months of flood peak (P 9 L10-17), local conditions are also lumped and lost as well. Therefore I suggest, rephrasing the entire paragraph and discussing the differences in the methods (how the flood seasons are defined) instead, with a focus on the outcomes of the flood season and not how the results are being applied to cluster regions in previous studies (which is only the second research step in most of the studies).

The authors agree, however we do stand by our claim that, compared with many studies, we offer results at a higher resolution. That said, these results may still not be considered 'local' in the sense that they are at the sub-basin scale. However this point, we believe, is minor. As the referee suggests, we have rephrased the text focusing on comparing methods and outcomes, and have eliminated text referring to the issue of clustering (also not imperative to make our case.) Please see the revised text in query 3 of major comments, above.

> P4 L27: Please specify what 'relaxing the criteria' means. What options have been assessed?

Relaxing the criteria refers to how the stations are selected. In this case, stations with records containing missing data are allowed, however even so there is no significant increase. We have removed the phrase altogether and changed P.4 L 27-28 to:

Although this criteria is admittedly quite strict (no missing 20-year daily data), including stations with missing records does not add a significant number. These stations are mostly located on large-rivers; annual streamflow of 75% of stations is larger than 100 m^3/sec, 35% of stations are larger than 500 m^3/sec, 20% of stations are larger than 1,000 m^3/sec and 5% of stations are larger than 5,000 m^3/sec.

> P 6 Section 3.1: The first paragraph that briefly reviews existing methods and explains the similarities and differences with the new approach, is still very confusing.
>
> Please re-write the section again in a more structured manner. Especially, please clarify what is meant by streamflow volume and magnitude (water level?)

The authors apologize for the lack of clarity. We have re-structured this section, as indicated below. Streamflow volume and magnitude have been replaced with peak volume and peak timing.

In the last few decades, a number of studies have investigated the timing of peak flows in the context of analyzing flood seasonality, frequency and trends. Generally, two main properties are emphasized regarding flood timing: peak volume and peak timing. Considering peak volume, the occurrence dates are commonly recorded for a fixed-time period or specific amount of peak volume, often in the context of trend analysis. For examples, Hodgkins and Dudley (2006) use winter-spring center of volume (WSCV) dates to analyze trends in snowmelt-induced floods, and Burn (2008) uses percentiles of annual streamflow volume dates as indicators of flood timing, also for trend analysis. For peak timing, two sampling methods are frequently applied in hydrology. The first and most common is the annual-maximum (AM) method, which samples the largest streamflow in each year. The second method is the peaks-over-threshold (POT) method (Smith, 1984, 1987; Todorovic and Zelenhasic, 1970), in which all distinct, independent dominant peak flows greater than a fixed threshold are counted, prior to a specified date. In contrast to the AM method, POT can capture multiple large independent floods within a single year, including the annual maximum flow, but may also miss the annual maximum flow in years in which streamflow is less than the pre-defined threshold (Cunderlik et al., 2004a.) The PM selected, therefore, is dependent on the peak properties (volume, timing) considered. For a local study, selecting the PM can be based on well-defined climatic or hydrologic characteristics (e.g. rainy season, snow-melt, etc.), however no single global method can be uniformly applied to define the PM everywhere. Thus, to define the HS, and specifically the PM, globally, both peak volume and peak timing aspects need to be considered (Javelle et al. 2003.) To do this, we adopt a Volume-Based Threshold (VBT) technique. This technique is similar to a streamflow volume-based technique in terms of capturing the days (Julian dates) when streamflow exceeds the pre-defined threshold (percentile of flows) and associated volume (Burn, 2008.) The major difference, however, is that the VBT applies the threshold over the entire time-series (available record) concurrently instead of on a year-by-year basis. In other words, for the 95th percentile, instead of annually calculating the 95th percentile, it is calculated using the entire period of record. The common volume-based technique thus records events every year surpassing the threshold, however for the VBT approach, every year need not have a peak above the threshold. This approach emphasizes capturing the key peaks across the entire available time-series (as in a peak over threshold approach.) VBT thus contains both volume and timing characteristics for defining the Peak Month (PM.) Here we select streamflow surpassing the top 5% of flows across all years (1958-2000) as the threshold for considering a high streamflow level; this level is commonly adopted in threshold approaches (Burn, 2008; Mishra et al., 2011.) The month containing the greatest number of occurrences in the top 5% is defined as the PM, and subsequently the HS is defined as the period containing the PM plus the month before and after the PM. Figure 2 provides an example based on seven years of synthetic streamflow; the number of days surpassing the 5% threshold is listed for each month. In this example, August has the largest number of days over the threshold (105 days), thus August is defined as PM and July-September is defined as HS.

P7 L 12: After reading the previous section it is still not clear why the PAMF 'inherently contains magnitude and volume properties'.

The authors apologize for the misunderstanding. To clarify, we have rephrase the sentence to indicate that the PAMF contains peak timing information, and have therefore changed P.7 L 12 to:

The PAMF is relatively simple, yet provides clear indication of how well PM selected represents the occurrence of annual peaks across the time-series.

> P7 L 20- P8 L2: The description of other methods should go a to a paragraph before the 'new method' is presented

The authors appreciate the suggestion. While we have not explicitly moved this paragraph up, we have introduced other methods (POT etc) prior to introducing the new methods, VBT. Additional description regarding the other methods is also provided simultaneous to the introduction of VBT.

> P8 L3-17: See 'General comment Number 3'. Cross correlation between different techniques can never 'indicate some success' of one method compared to the other! Correlation is not causation!

The authors agree. The purpose of using correlation is not to demonstrate the superiority of the VBT method, but rather to illustrate the similarity between the VBT method and all common volume-based methods. Please see the response to query 3 in major comments above for further details.

> P10 L 22-25: I suggest moving the discussion of Figure 7 further down in the document and discuss Figure 4, 5 and 6 first. Additionally, change 'United States and Canada', to 'North America' to be consistent with the labels in Figure 7.

The authors agree and have made the suggested changes.

> P 10 L27-30: It is mentioned that low PAMF values are computed for the US and Europe and this is attributed to be due 'at least in part, to reservoirs and dams along the Mississippi, Missouri and Danube'. This might true for the observed flows, however, with the modelled streamflow this should not be the case.
>
> However, from Figure 5 for example on can see that the model obtains even lower PAMF values in Europe! So certainly, the human impact does not play a role for the PAMF value to be low!

The referee makes a good point that we failed to explain. For parts of these regions, a relatively constant flow pattern may also cause the low PAMF values for modeled output. This is consistent with previous studies. Minor high-flow seasons may also play a role. Please see revised text a few comments below.

> As the analyses has a global focus, an in depth discussion (with more focus on spatial location) on the obtained differences in PM and PAMF is needed!

The authors have added discussion regarding additional locations. Please see revised text a few comments below.

> Additionally, for Europe and the US, only a few stations are actually located on the strongly anthropogenic impacted Rivers mentioned before. Therefore, the poor

performance shown with the PAMF might also be a shortcoming of the method and needs to be discussed further!

The authors agree and have added the suggested shortcomings of the method and outcomes. Please see revised text a few comments below.

P 11 L1-20: The discussion focusses too much on the areas where performance of the flood season can be considered acceptable. Areas with stronger differences such as Australia and South America are currently ignored and have to be discussed as well!

Additionally, it is highlighted that 40% of the models and the data share the same peak months. This is a quite low performance; however, the authors are not critical about this low outcome at all.

The authors agree and have added the suggested shortcomings of the method and outcomes. Please see revised text a few comments below.

Overall, I think in discussion the results, the authors should aim for a more balanced assessment of the good and less successful outcomes of their method!

The authors agree and have added the suggested shortcomings of the method and outcomes. Please see revised text below. P.10 L 21 – P.11 L 1 has been changed to:

In the southeastern United States, GRDC stations express relatively lower PAMF values for observations (40-60%) than model outputs (60-80%), due to the high level of managed infrastructure. In the central-southern United States and Europe, low PAMF values are computed for both observations and modeled output (Figure 5) with notable temporal differences (Figure 4 (c).) For observations, this is attributable, at least in part, to reservoirs and dams along the Mississippi, Missouri and Danube rivers. Additionally, relatively constant streamflow patterns are identified in both observations and modeled output, consistent with previous studies reporting these flow regimes as uniform or perpetually wet (Burn and Arnell, 1993; Dettinger and Diaz, 2000; Haines et al., 1988.) Minor high-flow seasons may also play a role. Model biases also effect PM selection; for Northwestern North America, PMs for many points are defined on average one month earlier than with observations, producing moderate PAMF values (60% and higher.) In Northern Europe, especially around Finland, this becomes much more pronounced, with large differences between PMs from observations and the model, on the order of 4-months (Figure 4(c), 6(c), and 8(a)). In western and northern Australia, PMs are modeled 1-month later on average than observations excepting with two occurrences in the west (5-month difference) due to both observed and modeled low-flow conditions. Such low-flow regimes are also apparent in southeastern Australia, causing large differences between PMs (4-5 months.) The differences in PMs between observations and modeled outputs are also compared at the continental scale (Figure 7.) In North America, 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 (e.g. HS; Figure 7.) In Asia 65% of stations and 70% of sub-basins have identical PMs, growing to 90% of stations and 92% of sub-basins with ±1 month temporal difference (Figure 7.) In central Russia, a large difference between PMs (± 3 months) are attributable to reservoirs on the Yenisei and Angara rivers and model bias (Figure 4 (c)). In Africa, 48% of stations and 60% of sub-basins produce identical PMs (Figure 7), 30% of stations and 27% of sub-basins are modeled 1-month earlier, and 7.4% of stations and 6.7% of sub-basins are modeled 1-month later than observation (Figure 7.) In South America, with only

5 stations, 40% have the same month, 40% are modeled 1-month earlier and 20% of stations are modeled 2-months earlier than observations.

Comparing observations and modeled output globally, 40% of the locations share the same PM. The model's bias is one of main reasons for this moderate performance; other important contributors include minor high-flow seasons, perpetually wet or dry regions, and anthropogenic effects such as reservoir regulation.

> P 12 L1: I strongly object the authors claim that there is a 'striking similarity' between the DFO and the modelled season and that this 'further supports the model's ability to appropriately identify the PM spatially'. Please rephrase.

The authors have toned down the claim, however we do believe there is noted similarity, especially when both the major and minor high flow seasons are considered. Thus we have changed P.11 L 32 - P.12 L 2 to:

Nevertheless, model-based PMs and historic flood records illustrate similarity (compare Figures 8 (a) and 9), particularly when both the major and minor high flow seasons are considered, further indicating merit in the ability of the proposed approach to identify the PM.

> P14 L27: Please rephrase 'streamflow magnitude and volume characteristics of floods' to something that explains the method better.

We have changed P.14 L 27-29 to:

In order to consider both peak volume and peak timing, a volume-based threshold technique is applied to define the high-flow season and is subsequently evaluated by the PAMF.

> P14 L 31-P15 L 2: Please rephrase, as there are only 40% of the peak months that are appropriately identified correctly, which is not an 'indication of strong agreement between model and observed flood season' and the flood records of the DFO are also not 'well represented', to be more realistic about the outcomes of the study.

The authors agree that while 40% may not be overly impressive, 89% within ±1 month is quite strong. This effectively implies that the model captures the high-flow season (3-month) for 89% of all selected locations. We have added more balance to this paragraph, changing P.14 L 30 – P.15 L 2 to:

As a result, 40% of stations and 50% of sub-basins have identical peak months and 81% of stations and 89% of sub-basins are within 1 month, thus well capturing high flow seasons. When considering anthropogenic effects and bi-modal or perpetually wet/dry flow regions, these results indicate fair agreement between modeled and observed high-flow seasons.

> P15 L 24-25: Please rephrase, as the model does not 'enable the complete flood season identification globally'. There are many locations on the globe where, there are problems and low performance as indicated by low PAMF values. Therefore, the method is not globally applicable. It would be better if it would be highlighted where the flood season can be expected to be well represented by the model!

To address we have changed P.15 L 22-25 to:

The global hydrological model's ability to define major and minor high-flow seasons at high resolution is highlighted in this study. Although results indicate relatively positive performance overall, regional performance varies spatially.

Figures:

Figure 8: I suggest merging Figure 8 and Figure 9 into one Figure with 2 panels to allow a better interpretation of the results. Having Figure 8 and 9 together helps to interpret the reliability of the months defined in Figure 8).

The authors agree and have merged Figure 8 and Figure 9.

Figure 9: From the current way of plotting, it is difficult to distinguish the different reliability classes as defined on page 7. For example, based on the classes defined beforehand, central Europe and most of Australia has a poor reliability. I suggest showing the PAMF not as gradual colours but actually show the colours according to the reliability classes defined beforehand so the results can be interpreted accordingly. This should also be better discussed in the text.

The authors agree and have changed the color code. Please see updated figures below.

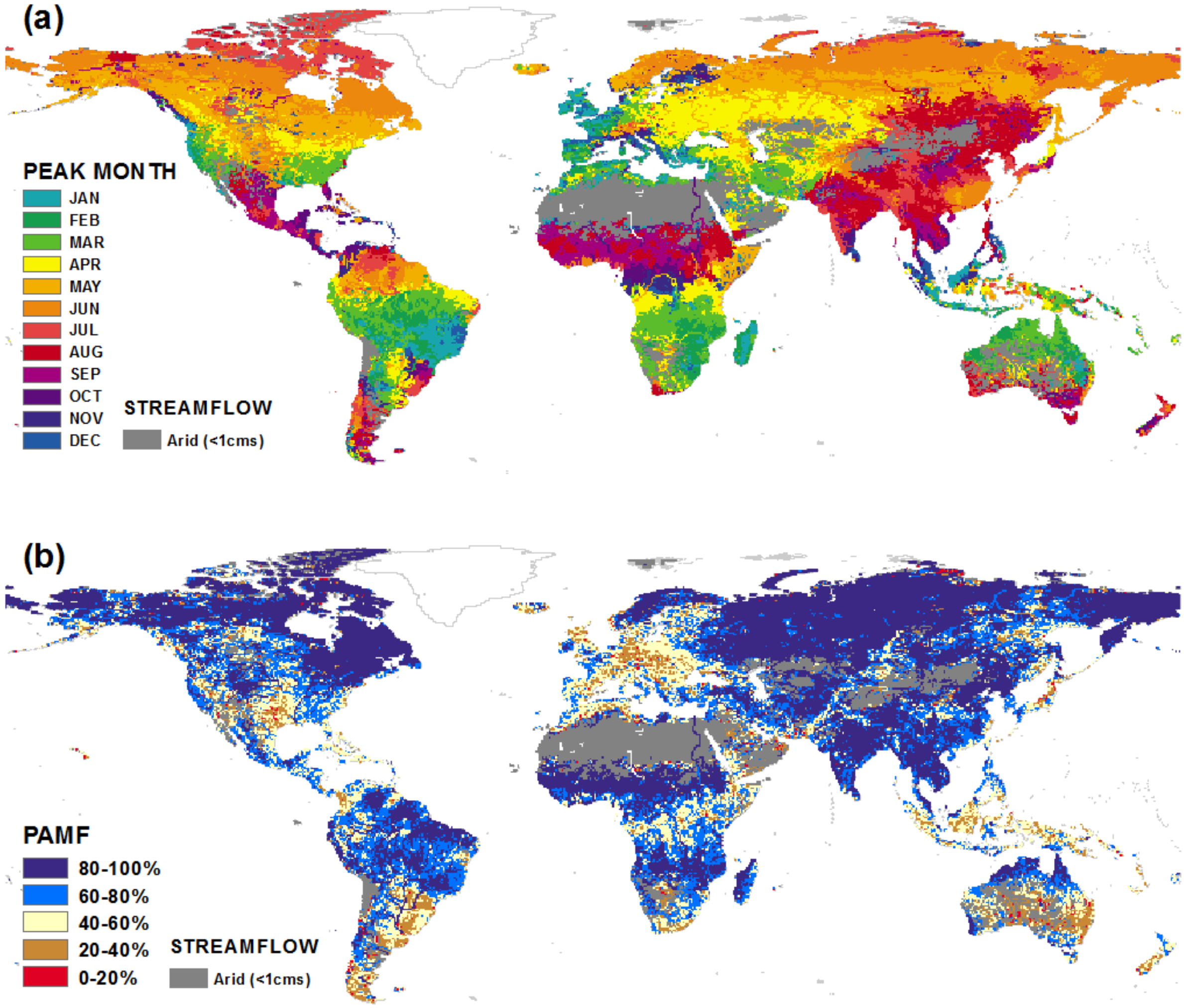


Figure 1. (a) Peak Month (PM) as defined at all modeled grid cells (b) Calculated Percentage of Annual Maximum Flow (PAMF) values for at all modeled grid cells; subjectively classified as high = 80-100%, moderate = 60-80%, low = 40-60% and poor = 0-40%.

Defining ~~Flood~~High-flow Seasons ~~Globally~~ using Temporal Streamflow Patterns from a Global Model

Donghoon Lee[1], Philip Ward[2] and Paul Block[1]

[1]{University of Wisconsin - Madison, Madison, Wisconsin, USA}

[2]{Institute for Environmental Studies (IVM), VU University Amsterdam, the Netherlands}

Corresponding Author: Paul Block (paul.block@wisc.edu)

Abstract

Globally, flood catastrophes lead all natural hazards in terms of impacts on society, causing billions of dollars of damages annually. Here, a novel approach to defining ~~flood~~high-flow seasons globally is presented by identifying temporal patterns of streamflow~~ objectively~~. The main ~~flood~~high-flow season is identified using a volume-based threshold technique and the PCR-GLOBWB model. In comparison with observations, 40% (50%) of locations at a station (sub-basin) scale have identical peak months and 81% (89%) are within 1 month, indicating ~~strong~~fair agreement between modeled and observed ~~flood~~high-flow seasons. ~~Model defined flood~~Minor high-flow seasons are ~~additionally~~also defined for bi-modal flow regimes. Identified major and minor high-flow seasons together are found to well represent actual flood records from the Dartmouth Flood Observatory, further substantiating the ~~models~~model's ability to reproduce the appropriate ~~flood~~high-flow season. ~~Minor flood~~These high-resolution high-flow seasons ~~are also defined~~and associated performance metrics allow for ~~bi-modal flood regimes~~an improve understanding of temporal characterization of streamflow and flood potential, causation, and management. This is especially attractive for regions with limited observations and/or little capacity to develop early warning flood systems.~~ The temporal patterns of global streamflow identified can lead to improved understanding of flood frequency, trends, and inter-annual variability and further benefit flood risk planning and preparation efforts.~~

1 Introduction

Flood disasters rank as one of the most destructive natural hazards in terms of economic damage, causing billions of dollars of damage each year (Munich Re, 2012.) These flood damages have risen starkly over the past half-century given the rapid increase in global exposure (Bouwer, 2011; UNISDR, 2011; Visser et al., 2014.) To specifically address flood disasters from a global perspective, understanding of global-scale flood processes and streamflow variability is important (Dettinger and Diaz, 2000; Ward et al., 2014). In recent decades, studies have investigated global-scale streamflow characteristics using observed streamflow from around the world (Beck et al., 2013; McMahon, 1992; McMahon et al., 2007; Peel et al., 2001, 2004; Poff et al., 2006; Probst and Tardy, 1987) and modeled streamflow from global hydrological models (Beck et al., 2015; van Dijk et al., 2013; McCabe and Wolock, 2008; Milly et al., 2005; Ward et al., 2013, 2014) to investigate ungauged and poorly gauged basins (Fekete and Vörösmarty, 2007). Despite this broad attention on annual streamflow and its connections to global climate processes and precursors, there has been relatively little attention paid to the intra-annual timing of streamflow, emphasizing the need for analysis of seasonal streamflow patterns to further improve understanding of large-scale hydrology and atmospheric behaviors on the main (flood) streamflow season globally (Dettinger and Diaz, 2000). Moreover, better assessment of streamflow timing and seasonality is important for addressing frequency and trend analyses, flood protection and preparedness, climate-related changes, and other hydrological applications that possess important sub-annual characteristics (Burn and Arnell, 1993; Burn and Hag Elnur, 2002; Cunderlik and Ouarda, 2009; Hodgkins et al., 2003). This motivates further investigation of intra-annual temporal streamflow patterns globally.

Only a small number of studies have investigated global-scale seasonality and temporal patterns of streamflow, with minimal focus on objective streamflow timing. Haines et al. (1988) cluster 969 world rivers into 15 categories based on seasonality and average monthly streamflow data, and present one of the first maps providing a global classification. Burn and Arnell (1993) aggregate 200 streamflow stations into 44 similar climatic regions and subsequently combine these into 13 groups using hierarchical clustering based on similarity of the annual maximum flow index, providing spatial and temporal coincidences of flood response. Dettinger and Diaz (2000) aggregate 1345 sites into 10 clusters based on seasonality using climatological fractional monthly flows (CFMFs) to identify peak months and linkages with large-scale climate drivers.

In general, these studies define ~~key~~high streamflow or flood seasons subjectively based on the relationship between dominant streamflow amplitude patterns and large-scale climate drivers/patterns~~ and clustering~~, and ~~describe~~delineate large-scale homogeneous regions correspondingly. Defining high flow season timing is essentially a bi-product of these analyses, and may be problematic due to varying seasonal patterns (e.g. bi-modal distribution, constant or low flow areas, etc.) not captured at the large-scale delineation. There is also typically no distinguishment between minor and high flow seasons. In some cases, these minor seasons (e.g. resulting from bi-modal precipitation distribution) can produce high flow or flood conditions, and are thus of interest to identify. Here we identify high-flow seasons by capturing annual peak timing ~~based~~using a volumetric technique at the cell and sub-basin scale, presenting an approach focused on streamflow temporal patterns rather than pattern of amplitude. The ~~clustering approach employed, however, tends to lump many smaller basins together into one major season, and may not be representative of local scale conditions, especially for those basins at the margins of the clusters where streamflow patterns may not have a single defined major~~new measure of PM (and HS) coupled with the model grid scale provides much higher resolution peak timings globally than previously presented (often at large basin scale or subcontinental scale.) The performance measure introduced here (PAMF) is also a new contribution relating the models ability to capture high flow season ~~(e.g. perpetually wet or dry, bi-modal, etc.) In lieu of large-scale clustering and aggregation, we propose addressing basins (and even grid cells) individually, in a disaggregated fashion. Additionally, as an alternative to simply using the~~timing. These advantages are also helpful for identifying less-dominant but important seasons (minor high-flow seasons) that possess similar characteristics to the high flow season (e.g. bi-modal annual ~~maximum streamflow amplitude to define the annual peak timing, we consider a volumetric approach based on surpassing a predefined threshold at the daily scale to define peak streamflow timing. Both of these advancements are conditioned on an objective approach to define major and minor flood seasons globally. Parsing out these peak seasons – if they exist – can improve flood~~ cycle), another unique contribution of this work. This leads to better temporal characterization~~, leading to better~~ and understanding of flood potential, causation, and management, particularly in ungauged or limited-gauged basins.

2 Data description

2.1 Streamflow stations

Daily streamflow observations utilized in this study are from the Global Runoff Data Centre ~~(GRDC, 2007)~~(GRDC, 2007), specifically those, stations located along the global hydrology model's drainage network. Since station records that are missing even short periods may effect how a ~~flood~~high-flow season is defined, we have excluded years with any daily missing values. In this study, a minimum of 20 hydrological years is required for a station to be retained, leaving, 691 stations from all continents except Antarctica, with upstream basin areas ranging from 9,539 to 4,680,000 km² and periods of record between 20 - 43 years across 1958 - 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~~20-year daily data), including stations with missing records does not add a significant number. These stations are mostly located on large-rivers; annual streamflow of 75% of stations is larger than 100 m^3/sec, 35% of stations are larger than 500 m^3/sec, 20% of stations are larger than 1,000 m^3/sec and 5% of stations are larger than 5,000 m^3/sec.

2.2 PCR-GLOBWB

In this study, we evaluate simulations of daily streamflow over the period 1958-2000 taken from Ward et al. (2013), carried out using PCR-GLOBWB (PCRaster GLOBal Water Balance), a global hydrological model with a 0.5° x 0.5° resolution (Van Beek and Bierkens, 2009; Van Beek et al., 2011.) Although the PCR-GLOBWB model is not calibrated, and simulations may contain biases and uncertainty at course spatial resolution, the long time-series of streamflow provided globally has been deemed sufficient to estimate long-term flow characteristics with spatial consistency (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), with strong model performance. Note that for the simulations used in this study, the maximum storage within the river channel is based on geomorphological laws that do not account for existing flood protection measures such as dikes and levees.

For the simulations used in this study, the PCR-GLOBWB model was forced with daily meteorological data from the WATCH (Water and Global Change) project (Weedon et al., 2011), namely precipitation, temperature, and global radiation data. These data are available at the same resolution as the hydrological model (0.5° x 0.5°.) 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, precipitation gauges, time-scale adjustments of daily values to reflect monthly observations, and varying atmospheric aerosol-loading. 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).

3 Defining ~~flood~~high-flow seasons

To identify spatial and temporal patterns of dominant streamflow uniformly, we design a fixed time window for representing ~~flood~~high-flow seasons globally. Here we define major ~~flood~~high-flow 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~~High-flow Season (~~FS~~HS.) Specifically, we define PM first, and then define ~~FS~~HS 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.

3.1 Methodology for defining grid-cell scale ~~flood~~high-flow seasons

In the last few decades, a number of studies have investigated the timing of ~~floods~~peak flows in the context of analyzing flood seasonality, frequency and trends. Generally, two main ~~factors~~properties are emphasized regarding flood timing: ~~streamflow~~peak~~ volume and~~ e~~streamflow magnitude. For streamflow~~peak timing. Considering peak~~ volume~~e~~ an~~, the occurrence ~~date is~~dates are commonly recorded for a fixed-time period or specific amount of peak volume, often in the context of trend analysis. For examples, Hodgkins and Dudley (2006) use winter-spring center of volume (WSCV) dates to analyze trends in snowmelt-induced floods, and Burn (2008) uses ~~percentile~~percentiles of annual streamflow volume dates as indicators of flood timing, also for trend analysis. ~~The second factor (streamflow magnitude) is traditionally more focused on~~For peak~~ flood~~ timing~~. Two~~, two sampling methods are frequently applied in hydrology. The first and most common is the annual-maximum (AM) method, which samples the largest streamflow in each year. The second method is the peaks-over-threshold (POT) method (Smith, 1984, 1987; Todorovic and Zelenhasic, 1970), in which all distinct, independent dominant peak flows greater than a fixed threshold are counted, prior to a specified date. In contrast to the AM method, this threshold characteristicPOT can capture multiple large independent floods within a single year, including the annual maximum flow, but may also miss the annual maximum flow in years in which streamflow is less than the pre-defined threshold (Cunderlik et al., 2004a). Thus, deciding the proper threshold level is important. Therefore, to define the FS, and specifically the PM, both volume and magnitude(Cunderlik et al., 2004a.) The PM selected, therefore, is dependent on the peak properties (volume, timing) considered. For a local study, selecting the PM can be based on well-defined climatic or hydrologic characteristics (e.g. rainy season, snow-melt, etc.), however no single global method can be uniformly applied to define the PM everywhere. Thus, to define the HS, and specifically the PM, globally, both peak volume and peak timing 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 occurredJavelle et al. 2003.) To do this, we adopt a Volume-Based Threshold (VBT) technique. This technique is similar to a streamflow volume-based technique in terms of capturing the days (Julian dates) when streamflow exceeds the pre-defined threshold (percentile of flows) and associated volume (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(Burn, 2008.) The major difference, however, is that the VBT applies the threshold over the entire time-series (available record) concurrently instead of on a year-by-year basis. In other words, for the 95^{th} percentile, instead of annually calculating the 95^{th} percentile, it is calculated using the entire period of record. The common volume-based technique thus records events every year surpassing the threshold, however for the VBT approach, every year need not have a peak above the threshold. This approach emphasizes capturing the key peaks across the entire available time-series (as in a peak over threshold approach.) VBT thus contains both volume and timing characteristics for defining the Peak Month (PM.) Here we select streamflow surpassing the top 5% of flows across all years (1958-2000) as the threshold for considering a high streamflow level; this level is commonly adopted in threshold approaches (Burn, 2008; Mishra et al., 2011.) The month containing the greatest number of occurrences in the top 5% is defined as the PM, and subsequently the FSHS is defined as the period containing the PM plus the month before and after the PM. Figure 2 provides an example based on seven years of synthetic streamflow; the number of days surpassing the 5% threshold is listed for each month. In this example, August has the largest number of days over the threshold (105 days), thus August is defined as PM and July-September is defined as ~~FS.~~ HS.

To evaluate the defined ~~FS~~HS objectively, by evaluating the number of annual maximum flows captured, we develop a simple evaluating statistic called the Percentage of Annual Maximum Flow (PAMF). PAMF is computed as shown in Eq. 1:

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

where $nAMF(i)$ denotes the number of annual maximum flows that occur in month i across the full record. 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~~HS across the evaluation period. The PAMF is relatively simple, yet ~~inherently contains magnitude and volume properties~~provides clear indication of ~~streamflow.~~how well PM selected represents the occurrence of annual peaks across the time-series. For example, a high PAMF indicates that the ~~FS~~HS 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~~HS PAMF values as: high = 80-100%, ~~low~~moderate = 60-80%, low = 40-60% and poor = 0-40~~-60~~%. The PAMF is calculated for both the observed streamflow at the selected 691 GRDC stations and the simulated streamflow at the associated 691 grid locations.

~~Clearly the volume-based threshold method is not the only available classification technique for defining the PM. To gauge its performance, the AM method and other volume methods with different given durations are selected for comparison, namely Q_{AM}, Q_{7day}, Q_{15day} and Q_{30day}. For the Q_{AM} approach, which is based on the AM method, the FS is simply centered on the PM containing the largest number of annual maximum flow occurrences across the total years available. The $Q_{7\mathrm{day}}$ approach defines the PM as the month with maximum streamflow volume during any seven consecutive day period; the month with the most periods across all years becomes the PM for the defined FS. The $Q_{15\mathrm{day}}$ and $Q_{30\mathrm{day}}$ approaches are similar to the $Q_{7\mathrm{day}}$ approach, respectively using 15 and 30 days consecutively. Comparitavely, 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. The volume-based threshold method is an attempt to bridge these two criteria.

The VBT technique is compared with the common volume-based technique and POT technique to gauge performance. Four volume-based durations, namely V01%, V03%, V05% and V10% and three POT techniques averaging 1, 2, and 3 peaks per year (POT1, POT2 and POT3, respectively) are selected. For the V01% technique, the HS is simply centered on the PM containing the largest number of occurrences of the top 1% of annual streamflow volume across the total years available. The V03%, V05% and V10% techniques are similar to the V01% approach, respectively using 3%, 5% and 10% of annual streamflow volume. Comparatively, techniques with a shorter time component (1-3% of annual volume) favor identifying the PM by peak timing while techniques with longer time components (5-10% of annual volume) favor identifying the PM based on duration and peak volume. The VBT technique is an attempt to bridge these two criteria. For the POT techniques, independence criteria is applied to avoid counting multiple peaks from the same event (Institute of Hydrology, 1999.) For example, two peaks must be separated by at least three-times the average rising time to peak, and minimum flow between two peaks must be less than two-thirds of the higher one of the two peaks. More details of independence criteria are described in Lang et al. (1999.)

An analysis examining sensitivity of selected threshold levels for the VBT technique is also undertaken. Performance of thresholds representing 1%, 3%, 5% and 10% exceedance across the entire period of record, named VBT1%, VBT3%, VBT5% and VBT10%, respectively, are compared.

To compare techniques and thresholds, the PMs are defined at the 691 selected stations and associated model grids. The locations where the PMs differ (by at least one technique) are of most interest. This occurs at 61% of stations and 54% of associated grids. Cross-correlations of PM between the four common volume-based techniques clearly indicate the tendency of the defined PM to shift from peak timing dominated to peak volume dominated as the time component increases (Table 1.) Correlation between VBT techniques and volume-based threshold technique and other classification techniques are quite similar (and consistent (0.82-0.86 and 0.84-0.86 for observed and simulated streamflow, using VBT5%;0.87-0.90; Table 1), preliminarily indicating some success in capturing both magnitude and volume. timing and volume properties, while correlation between the VBT techniques and POT are less strong

(0.78-0.81 and 0.79-0.83 for observed and simulated streamflow, respectively, using VBT5%; Table 1.) The PAMF is also useful for comparing ~~classification techniques' performance when they define PM differently at the same location. This occurs at 45% or stations and 40% of associated grids for observed and modeled streamflow, respectively. The classification~~techniques, such that the technique having the highest average PAMF ~~most often for those stations or girds may be considered slightly superior in terms of containing~~typically contains more annual maximum ~~flows~~flow events in their defined ~~FSs~~HSs. The ~~volume-based threshold technique has~~VBT5% is superior to other VBT and POT techniques for both observed and modeled streamflow, having the highest PAMF values~~ by at least 2% of grids (1% of stations) more than any other technique for modeled (observed) streamflow. ~~, however the volume-based techniques indicate similar or even slightly better performance than VBT5% (Table 2.) This is not unexpected as the volume-based techniques are designed to capture annual peak flows on a year-by-year basis, whereas the VBT records significant peaks across the full time-series, and may "miss" annual peaks in some years in which that peak is small relative to all peaks throughout the available record. Thus VBT tends to select PMs that contain the most significant peaks overall, and subsequently have the highest potential for capturing probable flood seasons for flood-prone basins, a desirable outcome for this study. To illustrate this in the context of the PAMF, if all years are ranked for each location based on the annual peak flow, and the top 50% (half) are retained, the PAMF actually favors the VBT approach, surpassing the volume-based approach by 5-6% for PMs and 2-3% for HSs.

Finally, ~~classification technique performance~~techniques may be evaluated by comparing the temporal difference (number of months) between model-based and observed PMs; closer is clearly superior. ~~Overall the volume-based threshold technique produces a greater~~ The VBT3% and VBT5% techniques produce the greatest degree of similarity between model-based and observed PMs (~~2-5% higher in~~ 81% of stations having ±1 month difference~~ and 1-5% higher in ±2 month difference.) Based on these findings~~; Table 3.) Overall, the VBT technique demonstrates superior performance as compared with the POT techniques by all comparisons. The VBT technique is also on par or slightly superior to the common volume-based technique, especially considering the 5% threshold; thus, the remainder of the analysis is carried out utilizing the ~~volume-based threshold~~VBT5% technique only.

3.2 Methodology for defining sub-basin scale ~~flood~~high-flow seasons

In addition to evaluating the ~~FS~~HS at the 691 grid cells based on model outputs, the PM and ~~FS~~HS can also be defined at the sub-basin scale globally where observations are present. Previous studies have investigated flood seasonality as it relates to basin characteristics; for example, basins are delineated/regionalized and grouped according to similarity/dissimilarity of streamflow seasonality (Burn, 1997; Cunderlik et al., 2004a), or conversely, flood seasonality is occasionally used to assess hydrological homogeneity of a group of regions (Cunderlik and Burn, 2002; Cunderlik et al., 2004b), thus evaluating at the sub-basin scale is warranted.

While defining a single PM for a large-scale basin may be convenient, it may be difficult to justify given the potentially long travel times and varying climate, topography, vegetation, etc. Additionally, infrastructure may be present to regulate flow for flood control, water supply, irrigation, recreation, navigation, and hydropower (WCD, 2000), causing managed and natural flow regimes to differ drastically. This becomes important, as globally more than 33,000 records of large dams and reservoirs are listed (ICOLD, 1998-2009), with geo-referencing available for 6,862 of them (Lehner et al., 2011). Nearly 50% of large rivers with average streamflow in excess of 1,000 m^3/s are significantly modulated by dams (Lehner et al., 2011), often significantly attenuating flow hydrographs and flood volumes~~.~~ (twenty percent of GRDC stations fall into this category.) The PAMF, as previously defined, can aid in identifying stations affected by upstream reservoirs through 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.) In this study, we used the global sub-basins from the 30' global drainage direction map (DDM30) dataset (Döll and Lehner, 2002) with separation of large basins (Ward et al., 2014).

To define a sub-basin's PM, the maximum PAMF and associated PM for each station within the sub-basin are considered according to the following:

- If multiple stations exist within the sub-basin, the PM is defined as the PM occurring for the largest number of stations
- If there is a tie between months, their average PAMF values are compared, and the month having the higher average PAMF is defined as the PM.
- If there is a tie between months and equivalent average PAMF values, the month having the higher average annual streamflow is defined as the PM.

The sub-basin’s PM is defined based on the occurrence of station or grid-level PMs rather than the PAMF values to diminish results being skewed by biased simulations or varying climate effects in small parts of the sub-basin. When there are an equal number of occurrences for different PMs, the average PAMF values are used to determine which PM is selected. In this case, the effect of stations downstream of reservoirs will be minimized given their typically low average PAMF values. This procedure is applied for both stations (observations) and corresponding grid cells (model) in each sub-basin. To illustrate, consider the 6 GRDC stations in the Zambezi River Basin (Figure 3.) For most of the stations, the observed PM is defined as a month later than the model-based PM (Table ~~2~~4), an apparent bias in the model. The PAMF of STA06 observations is noticeably lower than for other stations (36%; Table ~~2~~4) given its location downstream of the Itezhi-Tezhi dam (STA05) (Figure 3.) Otherwise, PAMF values are consistently high across all stations. March is the PM identified most often, thus the final sub-basin PM selected is March.

In contrast, the model-based simulated streamflow produces a high PAMF at STA06 (97%), as the Itezhi-Tezhi dam is not represented in the simulations used for this study, and subsequently does not account for modulated streamflow. Across other stations, the PAMF is also high, however an equal number of stations select February and March. In this case, February is selected as the final basin PM given its higher average PAMF value (96% vs. 91%.)

By this approach, all 691 GRDC stations are grouped into 223 sub-basins to define the PM (Figure 6.); 58% of sub-basins are defined by a single station, only 7.6% (observations) and 8.1% (model) of sub-basins have ties when defining PMs, and only one sub-basin has a tie between PMs and average PAMF values.

4 Verification of selected ~~flood~~high-flow seasons

Model-based PMs are verified by comparing with observation-based PMs at station and sub-basin scales. Additionally, historic flood records from the Dartmouth Flood Observatory (DFO) are used to compare basin-level PMs to actual flooded areas spatially and temporally. Specifically, we apply the following information from DFO: start time, end time, duration and geographically estimated area at 3,486 flood records across 1985-2008.

4.1 Observed versus modeled ~~flood~~high-flow seasons

~~Ideally the model-based and observed GRDC stations have fully or partially overlapping FS periods. If so, this builds confidence in interpreting FSs at locations where no observed data are available. 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). These temporal differences are also compared at the continental scale (Figure 7.) 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 (e.g. FS; Figure 7.) GRDC stations in the southeastern United States express relatively lower PAMF values for observations (40-60%) than model outputs (60-80%), due to the high level of managed infrastructure. In the central United States and Europe, low PAMF values are computed for both observation and model outputs (Figure 5) with notable temporal differences (Figure 4 (c).) This is attributable, at least in part, to reservoirs and dams along the Mississippi, Missouri and Danube rivers.~~

~~Globally, comparing model and GRDC data, 40% of the locations share the same PM.~~Ideally the model-based and observed GRDC stations have fully or partially overlapping HS periods. If so, this builds confidence in interpreting HSs at locations where no observed data are available. 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). In the southeastern United States, GRDC stations express relatively lower PAMF values for observations (40-60%) than model outputs (60-80%), due to the high level of managed infrastructure. In the central-southern United States and Europe, low PAMF values are computed for both observations and modeled output (Figure 5) with notable temporal differences (Figure 4 (c).) For observations, this is attributable, at least in part, to reservoirs and dams along the Mississippi, Missouri and Danube rivers. Additionally, relatively constant streamflow patterns are identified in both observations and modeled output, consistent with previous studies reporting these flow regimes as uniform or perpetually wet (Burn and Arnell, 1993; Dettinger and Diaz, 2000; Haines et al., 1988.) Minor high-flow seasons may also play a role. Model biases also effect PM selection; for Northwestern North America, PMs for many points are defined on average one month earlier than with observations, producing moderate PAMF values (60% and higher.) In Northern

Europe, especially around Finland, this becomes much more pronounced, with large differences between PMs from observations and the model, on the order of 4-months (Figure 4(c), 6(c), and 8(a)). In western and northern Australia, PMs are modeled 1-month later on average than observations excepting with two occurrences in the west (5-month difference) due to both observed and modeled low-flow conditions. Such low-flow regimes are also apparent in southeastern Australia, causing large differences between PMs (4-5 months.) The differences in PMs between observations and modeled outputs are also compared at the continental scale (Figure 7.) In North America, 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 (e.g. HS; Figure 7.) In Asia 65% of stations and 70% of sub-basins have identical PMs, growing to 90% of stations and 92% of sub-basins with ±1 month temporal difference (Figure 7.) In central Russia, a large difference between PMs (± 3 months) are attributable to reservoirs on the Yenisei and Angara rivers and model bias (Figure 4 (c)). In Africa, 48% of stations and 60% of sub-basins produce identical PMs (Figure 7), 30% of stations and 27% of sub-basins are modeled 1-month earlier, and 7.4% of stations and 6.7% of sub-basins are modeled 1-month later than observation (Figure 7.) In South America, with only 5 stations, 40% have the same month, 40% are modeled 1-month earlier and 20% of stations are modeled 2-months earlier than observations.

Comparing observations and modeled output globally, 40% of the locations share the same PM. The model's bias is one of main reasons for this moderate performance; other important contributors include minor high-flow seasons, perpetually wet or dry regions, and anthropogenic effects such as reservoir regulation. Considering a difference of ± 1 month, this jumps to 81%, and 91% for ± 2 months (Figure 7.) From a sub-basin perspective, the similarities are even stronger (50% identical PM, 88% ± 1 month and 92% ± 2 month), indicating a relatively high level of agreement. For locations having dissimilar PMs (≥ ± 3 months, 9% of locations and 8% of sub-basins), a substantial portion are located downstream of reservoirs directly, such as STA06 in the Zambezi example (Table ~~1~~4), or are low-flow (dry) or constant-flow locations, both producing exceedingly low PAMF values. Differences in PMs are not unexpected for low-flow and constant-flow locations, given the propensity for the annual streamflow maximum to potentially occur in a wide number of months. Overall, however, as more than 80% of both stations and sub-basins have similar PMs (± 1 month), it appears that the global water balance model performs appropriately well in defining ~~flood~~high-flow seasons globally at locations where observations are available.

This may be subsequently extended to defining PMs and PAMF at all grid cells (~~Figures~~Figure 8~~-9~~.) Generally, ~~a~~ low and poor PAMF ~~indicates an~~values (0-60%) indicate a naturally unstable annual maximum flow~~,~~ (no clear high-flow season), which occurs in cases of constant-flow, low-flow, bi-modal flow and regulated flow. All cases, except regulated flow, are simulated within the PCR-GLOBWB simulations used, thus the cell-based PAMF values (Figure ~~9)~~8 (b)) can provide a sense of confidence for the defined PM (Figure 8~~.)~~ (a).) Examples of low-flow regions include the central United States and Australia having low PAMF regional values (Figure ~~9.)~~8 (b).) Bi-modal regions, such as much of East Africa with its two rainy seasons, may also be associated with low PAMF values.

4.2 Modeled ~~flood~~high-flow seasons versus actual flood records

Model-based PMs may also be verified (subjectively) by surveying historic flood records. One such source is the Dartmouth Flood Observatory (DFO), a large, publically accessible repository of major flood events globally over 1985-2008, based on media and governmental reports and instrumental and remote sensing sources. Delineations of affected areas are best estimates (Brakenridge, 2011.) The DFO records provide duration of each flooding event, as defined by the report or source, and represented as occurrence month (Figure ~~10~~9.) DFO flood events and grid cell based PMs (Figure 8~~)~~ (a)) may be compared outright, however their characteristics differ slightly. The DFO covers 1985-2008 while the model represents 1958-2000. Also, the model-based PM represents the month most likely for a flood to occur; the DFO is simply a reporting of when the event did occur, regardless of whether it fell in the expected ~~flood~~high-flow season or not. Nevertheless, model-based PMs and historic ~~flooding~~flood records illustrate ~~a striking~~ similarity (compare Figures 8 (a) and ~~10),~~9), particularly when both the major and minor high flow seasons are considered, further ~~supporting~~indicating merit in the ~~model's~~ ability of the proposed approach to ~~appropriately~~ identify the PM~~ spatially~~. Consistently, regions with high model-based PAMF (80-100%), such as Eastern South America, Central Africa and Central Asia, tend to agree well with DFO records, while poor or less than poor PAMF (0-60%) regions, such as Central North America, Europe, and East Africa, tend not to be in agreement with DFO records. In these low PAMF regions, however, DFO records also illustrate floods occurring sporadically throughout the year, further supporting accordance between cell-based PAMF and DFO records (Figures ~~9~~8 (b) and ~~10~~9.)

5 Defining minor ~~flood~~high-flow seasons

In some climatic regions, there is no one single, well-defined flood season. For example, East Africa has two rainy seasons, the major season from June to September and the minor season from January to April/May. These two seasons are induced by northward and southward shifts of the inter-tropical convergence zone (ITCZ) (Seleshi and Zanke, 2004.) This bi-modal East African pattern allows for potential flooding in either season. In Canada, as another example, the dominant spring snowmelt season (Mar-May) and fall rainy season (Aug-Oct) allow for flood occurrences in either period (Cunderlik and Ouarda, 2009.)

Previous studies have investigated techniques to differentiate seasonality from uni-, bi- and multi-modal streamflow climatologies and evaluate trends in timing and magnitude of streamflow, including the POT method, directional statistics method, and relative flood frequency method (Cunderlik and Ouarda, 2009; Cunderlik et al., 2004a). These methods may perform well at the local (case-specific) scale to define minor ~~flood~~high-flow seasons, however applying them uniformly at the global scale can be problematic, given spatial heterogeneity. Additionally, even though bimodal streamflow climatology may be detected, the magnitude of streamflow in the minor season may or may not be negligible in regards to flooding potential as compared with the major season.

To detect noteworthy minor ~~flood~~high-flow seasons globally, we classify streamflow regimes by climatology and monthly PAMF value, calculated using Eq. (1) at each month (Figure ~~11~~10.) 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 (and may therefore contain a minor season); both constant and low-flow classifications represent low values of PAMF between months. Distinguishing between bi-modal and other classifications is nontrivial. For example, initial inspection of the constant streamflow classification (both climatology and monthly PAMF, Figure ~~11~~10 (c)) could be mistaken for a non-dominant bi-modal distribution. 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 m^3/sec.
- The major and minor PMs must be separated by at least two months in order to prevent an overlap of each ~~FS~~HS (3-month.)
- If there is a peak in the monthly PAMF values outside the major ~~FS~~HS, it is regarded as a *potential* minor PM. If the sum of the major and *potential* minor PM's PAMF is greater

than 60% (minimum of 29 out of 43 annual maximums fall in one of the ~~FS~~HS), the *potential* minor PM is confirmed as a minor PM; the major PM's PAMF cannot exceed 80%.

A *potential* minor PM is identified by a secondary peak in the monthly PAMF rather than the magnitude or shape of streamflow. A minor ~~FS~~HS 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~~10 (a)). The sum of both major and minor PM's PAMF (joint PAMF) is used to determine the likelihood that one of the ~~FSs~~HSs contains the annual maximum flow; a high value of the joint PAMFs (80-100%) indicates strong likelihood (Figure ~~11~~10 (b)), moderate values (60-80%) imply moderate likelihood, with some probability of being classified as constant streamflow (Figure ~~11~~10 (c)); low values (40-60%) are likely constant or low streamflow (Figure ~~11~~10 (d)). Minor ~~FSs~~HSs are similar to major ~~FSs~~HSs, containing the minor PM and the month before and after. Minor ~~FSs~~HSs are evident in the tropics and sub-tropics and are spatially consistent with bi-modal rainfall regimes discovered by Wang (1994) (Figure ~~12~~11.) Examples include East Africa (second rainy season in winter) and Canada (rainfall-dominated runoff in autumn) both having high joint PAMF values (80-100%.) Additional examples include the major ~~FS~~HS (NDJ) and minor ~~FS~~HS (MAM) in Central Africa consistent with the latitudinal movement of the ITCZ, intra-Americas' major ~~FS~~HS (ASON) and minor ~~FS~~HS (AMJJ) (Chen and Taylor, 2002), and coastal regions of British Columbia in Canada and southern Alaska's minor ~~FS~~HS (SOND) due to wintertime migration of the Aleutian low from the central north Pacific (Figure ~~12~~11.) Distinct runoff process controlled by different climate and hydrology 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 the major ~~FS~~HS (AMJ) is dominated by snowmelt and the minor ~~FS~~HS (JAS) is spurred on by the Asian monsoon. The same mechanism produces minor ~~FSs~~HSs around the extents of the Asian summer monsoon (90-100% of sum of PAMFs) (Figure ~~9~~8 (b) and ~~12~~11.) Moderate minor ~~FSs~~HSs include, for example, the southern United States' (Texas and Oklahoma) bi-modal rainfall pattern (AMJ and SON) and in the southwestern United States (Arizona) where the summertime major ~~FS~~HS (JJA) is produced by the North American monsoon and the wintertime minor ~~FS~~HS (DJF) is affected by the regional large-scale low pressure system (Woodhouse, 1997). Southeastern Brazil's summertime major ~~FS~~HS (NDJF) and post-summer minor ~~FS~~HS (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~~HS (FMAM) and minor ~~FS~~HS (JJA) are defined as moderate (60-80% of joint PAMF values for central Europe and 70%-90% for eastern Europe), indicating that a minor ~~FS~~HS is not overly pronounced; for northeastern Europe the major ~~FS~~HS (MAM) and minor ~~FS~~HS (NDJ) contain high joint PAMF values (80%-100%.)

For the major ~~FS~~HS and minor ~~FS~~HS with joint PAMF values exceeding 60% (Figure ~~13~~12), flood records (DFO) occurring over more than one month are counted in each month based on the reported duration. Although one distinct flood event may dominate a monthly DFO record, strong similarity is evident between the ~~FSs~~HSs and monthly flood records (Figure ~~13~~12.) Minor ~~FSs~~HSs with high PAMF values corresponding well with observed DFO flood records include East Africa (bi-modal streamflow), the intra-Americas, and Northern Asia; only a few reported flood records occur in the minor ~~FSs~~HSs in high latitudes.

6 Conclusions and Discussion

In this study, a novel approach to defining ~~flood~~high-flow seasons globally is presented by identifying temporal patterns of streamflow objectively. Simulations of daily streamflow from the PCR-GLOBWB model are evaluated to define the dominant and minor ~~flood~~high-flow seasons globally. In order to consider both ~~streamflow magnitude and~~ peak volume ~~characteristics of floods~~and peak timing, a volume-based threshold technique is applied to define the ~~flood~~high-flow season and is subsequently evaluated by the PAMF. To verify model defined ~~flood~~high-flow seasons, we compare with observations at both station and sub-basin scales. As a result, 40% of stations and 50% of sub-basins have identical peak months and 81% of stations and 89% of sub-basins are within 1 month, ~~indicating strong~~thus well capturing high flow seasons. When considering anthropogenic effects and bi-modal or perpetually wet/dry flow regions, these results indicate fair agreement between ~~model~~modeled and observed ~~flood seasons.~~ high-flow seasons. Regions expressing bi-modal streamflow climatology are also defined to illustrate potential for noteworthy secondary (minor) high-flow seasons. Model defined ~~flood~~major and minor high-flow seasons are additionally found to ~~well~~ represent actual flood records from the Dartmouth Flood Observatory, further substantiating the models ability to reproduce the appropriate ~~flood season. Regions expressing bi-modal streamflow climatology are also defined to illustrate potential for noteworthy secondary (minor) flood seasons~~high-flow season.

Large-scale temporal phenomena associated with the defined major and minor ~~flood~~high-flow seasons are also identified. For example, global monsoon systems are clearly evident, as driven by the ITCZ, in central and eastern Africa, Asia and northern South America (Figure 8.) Latitudinal patterns in the extra-tropics are also quite distinct, with ~~flood~~high-flow seasons often occurring across similar months in the year. These broad temporal patterns are consistent with previous findings ~~(e.g. Burn and Arnell, 1993; Dettinger and Diaz, 2000; Haines et al., 1988)~~(e.g. Burn and Arnell, 1993; Dettinger and Diaz, 2000; Haines et al., 1988), however this analysis goes further by not being constrained to large-scale patterns for seasonal definition (via clustering) and also providing a sense of the reliability of the defined ~~flood~~high-flow seasons. Specifically, the defined PM (Figure 8~~)~~ (a)) has extended Dettinger and Diaz (2000)'s Peak Months by focusing on basin and grid scale streamflow volumes and providing likelihood type maps using the PMAF metric developed here (e.g. Figure ~~9)~~8 (b)) to represent the reliability of the defined PM. This can provide a clear sense of whether the identified ~~flood~~high-flow season is pronounced or vague. ~~-~~The identification of minor ~~flood~~high-flow seasons and deciphering bi-modal from constant streamflow regimes is another notable contribution of this study; minor seasons have not been well identified in previous studies. These identified ~~flood~~high-flow seasons are also consistent with DFO flood records both spatially and temporally, further substantiating their appropriateness.

~~Although biased simulations may theoretically contribute to a misidentified flood season, the global hydrological model's ability to well define flood seasons is highlighted in this study. The full global coverage of streamflow data in the model enables complete flood season identification globally. This is advantageous for many reasons, including hydrologic assessment in ungauged and poorly gauged basins and also for investigating flood season timing within large basins having diverse physical processes, for example, how the PM may shift along long rivers (e.g. Congo River) or basins with both snowmelt and rain-dominated processes. These spatially heterogeneous flood seasons at high resolution have the potential to better characterize streamflow regimes than previous studies (e.g. Dettinger and Diaz, 2000; Haines et al., 1988). Additional analysis to include upstream management and regulations is required to further classify global streamflow regimes and major flood seasons (or the elimination of them) for specific subbasin-level hydrologic applications.~~

Although biased simulations may theoretically contribute to a misidentified high-flow season, the global hydrological model's acceptable ability to define high-flow seasons is highlighted in this study. The global hydrological model's ability to define major and minor high-flow seasons at high resolution is highlighted in this study. Although results indicate relatively positive performance overall, regional performance varies spatially. This is advantageous for many reasons, including hydrologic assessment in ungauged and poorly gauged basins and also for investigating flood season timing within large basins having diverse physical processes, for example, how the PM may shift along long rivers (e.g. Congo River) or basins with both snowmelt and rain-dominated processes. These spatially heterogeneous high-flow seasons at high resolution have the potential to better characterize streamflow regimes than previous studies (e.g. Dettinger and Diaz, 2000; Haines et al., 1988). Additional analysis to include upstream management and regulations is required to further classify global streamflow regimes and major high-flow seasons (or the elimination of them) for specific subbasin-level hydrologic applications.

Acknowledgements. Philip WardThe first author was partially funded by a grant from the University of Wisconsin-Madison. The second author was funded by a VENI grant from the Netherlands Organisation for Scientific Research (NWO).

References

Beck, H. E., van Dijk, A. I. J. M., Miralles, D. G., de Jeu, R. A. M., Sampurno Bruijnzeel, L. A., McVicar, T. R. and Schellekens, J.: Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, Water Resour. Res., 49(12), 7843–7863, doi:10.1002/2013WR013918, 2013.

Beck, H. E., de Roo, A. and van Dijk, A. I. J. M.: Global ~~maps~~Maps of ~~streamflow characteristics based~~Streamflow Characteristics Based on ~~observations~~Observations from ~~several thousand catchments,~~Several Thousand Catchments*, J. Hydrometeorol., ~~150423121816007~~16(4), 1478–1501, doi:10.1175/JHM-D-14-0155.1, 2015.

Van Beek, L. P. H. and Bierkens, M. F. P.: The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification, Utrecht., 2009.

Van Beek, L. P. H., Wada, Y. and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47(7), W07517, doi:10.1029/2010WR009791, 2011.

Bouwer, L. M.: Have Disaster Losses Increased Due to Anthropogenic Climate Change?, Bull. Am. Meteorol. Soc., 92(1), 39–46, doi:10.1175/2010BAMS3092.1, 2011.

Brakenridge, G. R.: Global Active Archive of Large Flood Events, Dartmouth Flood Observatory, University of Colorado. [online] Available from: http://floodobservatory.colorado.edu/Archives/index.html, 2011.

Burn, D. H.: Catchment similarity for regional flood frequency analysis using seasonality measures, J. Hydrol., 202(1-4), 212–230, doi:10.1016/S0022-1694(97)00068-1, 1997.

Burn, D. H.: Climatic influences on streamflow timing in the headwaters of the Mackenzie River Basin, J. Hydrol., 352, 225–238, doi:10.1016/j.jhydrol.2008.01.019, 2008.

Burn, D. H. and Arnell, N. W.: Synchronicity in global flood responses, J. Hydrol., 144(1-4), 381–404, doi:10.1016/0022-1694(93)90181-8, 1993.

Burn, D. H. and Hag Elnur, M. a.: Detection of hydrologic trends and variability, J. Hydrol., 255(1-4), 107–122, doi:10.1016/S0022-1694(01)00514-5, 2002.

Chen, a. A. and Taylor, M. a.: Investigating the link between early season Caribbean rainfall and the El Nino+1 year, Int. J. Climatol., 22(1), 87–106, doi:10.1002/joc.711, 2002.

Cunderlik, J. M. and Burn, D. H.: Local and Regional Trends in Monthly Maximum Flows in Southern British Columbia, Can. Water Resour. J., 27(2), 191–212, doi:10.4296/cwrj2702191, 2002.

Cunderlik, J. M. and Ouarda, T. B. M. J.: Trends in the timing and magnitude of floods in Canada, J. Hydrol., 375(3-4), 471–480, doi:10.1016/j.jhydrol.2009.06.050, 2009.

Cunderlik, J. M., Ouarda, T. B. M. J. and Bobée, B.: Determination of flood seasonality from hydrological records / Détermination de la saisonnalité des crues à partir de séries hydrologiques, Hydrol. Sci. J., 49(3), 511–526, doi:10.1623/hysj.49.3.511.54351, 2004a.

Cunderlik, J. M., Ouarda, T. B. M. J. and Bobée, B.: On the objective identification of flood seasons, Water Resour. Res., 40(1), n/a–n/a, doi:10.1029/2003WR002295, 2004b.

Dettinger, M. D. and Diaz, H. F.: Global Characteristics of Stream Flow Seasonality and Variability, J. Hydrometeorol., 1(4), 289–310, doi:10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2, 2000.

Van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J. and Beck, H. E.: Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide, Water Resour. Res., 49(5), 2729–2746, doi:10.1002/wrcr.20251, 2013.

Döll, P. and Lehner, B.: Validation of a new global 30-min drainage direction map, J. Hydrol., 258(1-4), 214–231, doi:10.1016/S0022-1694(01)00565-0, 2002.

Fekete, B. M. and Vörösmarty, C. J.: The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements, Predict. Ungauged Basins PUB Kick-off (Proceedings PUB Kick-off Meet. held Bras. Novemb. 2002), IAHS Publ. no. 309, 129–136, 2007.

Global Runoff Data Centre: Major River Basins of the World / Global Runoff Data Centre, Koblenz, Germany: Federal Institute of Hydrology (BfG). [online] Available from: http://grdc.bafg.de, 2007.

Guha-Sapir, D., Hoyois, P. and Below, R.: Annual Disaster Statistical Review 2013: The Numbers and Trends, Brussels: CRED., 2014.

Haines, a. T., Finlayson, B. L. and McMahon, T. a.: A global classification of river regimes, Appl. Geogr., 8(4), 255–272, doi:10.1016/0143-6228(88)90035-5, 1988.

Herdies, D. L.: Moisture budget of the bimodal pattern of the summer circulation over South America, J. Geophys. Res., 107(D20), 8075, doi:10.1029/2001JD000997, 2002.

Hodgkins, G. ., Dudley, R. . and Huntington, T. .: Changes in the timing of high river flows in New England over the 20th Century, J. Hydrol., 278(1-4), 244–252, doi:10.1016/S0022-1694(03)00155-0, 2003.

Hodgkins, G. a. and Dudley, R. W.: Changes in the timing of winter-spring streamflows in eastern North America, 1913-2002, Geophys. Res. Lett., 33, 1–5, doi:10.1029/2005GL025593, 2006.

ICOLD (International Commission on Large Dams): World Register of Dams. Version updates 1998-2009, Paris: ICOLD. [online] Available from: www.icold-cigb.net, 2009.

Institute of Hydrology: Flood Estimation Handbook, vol. 3., Institute of Hydrology, Wallingford, UK., 1999.

Javelle, P., Ouarda, T. B. M. J. and Bobée, B.: Spring flood analysis using the flood-duration-frequency approach: application to the provinces of Quebec and Ontario, Canada, Hydrol. Process., 17(18), 3717–3736, doi:10.1002/hyp.1349, 2003.

Lang, M., Ouarda, T. B. M. J. and Bobée, B.: Towards operational guidelines for over-threshold modeling, J. Hydrol., 225(3-4), 103–117, doi:10.1016/S0022-1694(99)00167-5, 1999.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N. and Wisser, D.: High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9(9), 494–502, doi:10.1890/100125, 2011.

Lima, K. C. and Satyamurty, P.: Post-summer heavy rainfall events in Southeast Brazil associated with South Atlantic Convergence Zone, Atmos. Sci. Lett., 11(1), n/a–n/a, doi:10.1002/asl.246, 2010.

McCabe, G. J. and Wolock, D. M.: Joint Variability of Global Runoff and Global Sea Surface Temperatures, J. Hydrometeorol., 9(4), 816–824, doi:10.1175/2008JHM943.1, 2008.

McMahon, T. A.: Global runoff: continental comparisons of annual flows and peak discharges, Catena Verlag, Cremlingen-Destedt, Germany., 1992.

McMahon, T. a., Vogel, R. M., Peel, M. C. and Pegram, G. G. S.: Global streamflows – Part 1: Characteristics of annual streamflows, J. Hydrol., 347(3-4), 243–259, doi:10.1016/j.jhydrol.2007.09.002, 2007.

Milly, P. C. D., Dunne, K. a and Vecchia, a V: Global pattern of trends in streamflow and water availability in a changing climate, Nature, 438(7066), 347–350, doi:10.1038/nature04312, 2005.

Mishra, A. K., Singh, V. P. and Özger, M.: Seasonal streamflow extremes in Texas river basins: Uncertainty, trends, and teleconnections, J. Geophys. Res. Atmos., 116(8), 1–28, doi:10.1029/2010JD014597, 2011.

Munich Re: Topics Geo 2012 Issue. Natural Catastrophes 2011. Analyses, Assessments, Positions, Münchener Rückversicherungs-Gesellschaft, Munich., 2012.

Peel, M. C., McMahon, T. a. and Finlayson, B. L.: Continental differences in the variability of annual runoff-update and reassessment, J. Hydrol., 295(1-4), 185–197, doi:10.1016/j.jhydrol.2004.03.004, 2004.

Peel, M. C., McMahon, T. A., Finlayson, B. L. and Watson, F. G. R.: Identification and explanation of continental differences in the variability of annual runoff, J. Hydrol., 250(1-4), 224–240, doi:10.1016/S0022-1694(01)00438-3, 2001.

Poff, N. L., Olden, J. D., Pepin, D. M. and Bledsoe, B. P.: Placing global stream flow variability in geographic and geomorphic contexts, River Res. Appl., 22(2), 149–166, doi:10.1002/rra.902, 2006.

Probst, J. L. and Tardy, Y.: Long range streamflow and world continental runoff fluctuations since the beginning of this century, J. Hydrol., 94(3-4), 289–311, doi:10.1016/0022-1694(87)90057-6, 1987.

Seleshi, Y. and Zanke, U.: Recent changes in rainfall and rainy days in Ethiopia, Int. J. Climatol., 24(8), 973–983, doi:10.1002/joc.1052, 2004.

Smith, R. L.: Threshold Methods for Sample Extremes, in Statistical Extremes and Applications, pp. 621–638, D. Reidel, Dordrech, Netherlands., 1984.

Smith, R. L.: Estimating Tails of Probability Distributions, Ann. Stat., 15(3), 1174–1207, doi:10.1214/aos/1176350499, 1987.

Todorovic, P. and Zelenhasic, E.: A Stochastic Model for Flood Analysis, Water Resour. Res., 6(6), 1641–1648, doi:10.1029/WR006i006p01641, 1970.

UNISDR: Global Assessment Report on Disaster Risk Reduction (GAR 11) - Revealing Risk, Redefining Development, Geneva, Switzerland: United Nations Office for Disaster Risk Reduction (UNISDR)., 2011.

Uppala, S. M., KÅllberg, P. W., Simmons, a. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. a., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. a., Beljaars, a. C. M., Berg, L. Van De, Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. a. E. M., Jenne, R., Mcnally, a. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. a., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J.: The ERA-40 re-analysis, Q. J. R. Meteorol. Soc., 131(612), 2961–3012, doi:10.1256/qj.04.176, 2005.

Visser, H., Petersen, A. C. and Ligtvoet, W.: On the relation between weather-related disaster impacts, vulnerability and climate change, Clim. Change, 461–477, doi:10.1007/s10584-014-1179-z, 2014.

Wada, Y., van Beek, L. P. H., Viviroli, D., Dürr, H. H., Weingartner, R. and Bierkens, M. F. P.: Global monthly water stress: 2. Water demand and severity of water stress, Water Resour. Res., 47(7), n/a–n/a, doi:10.1029/2010WR009792, 2011.

Wang, B.: Climatic Regimes of Tropical Convection and Rainfall, J. Clim., 7(7), 1109–1118, doi:10.1175/1520-0442(1994)007<1109:CROTCA>2.0.CO;2, 1994.

Ward, P. J., Eisner, S., Flörke, M., Dettinger, M. D. and Kummu, M.: Annual flood sensitivities to El Niño–Southern Oscillation at the global scale, Hydrol. Earth Syst. Sci., 18(1), 47–66, doi:10.5194/hess-18-47-2014, 2014.

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., van Beek, R., Bierkens, M. F. P., Ligtvoet, W. and Winsemius, H. C.: Assessing flood risk at the global scale: Model setup, results, and sensitivity, Environ. Res. Lett., 8(4), 44019, doi:10.1088/1748-9326/8/4/044019, 2013.

WCD (World Commission on Dams): Dams and development: a framework for decision making, London, UK: Earthscan., 2000.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O. and Best, M.: Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, J. Hydrometeorol., 12(5), 823–848, doi:10.1175/2011JHM1369.1, 2011.

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J. and Bouwman, a.: A framework for global river flood risk assessments, Hydrol. Earth Syst. Sci., 17(5), 1871–1892, doi:10.5194/hess-17-1871-2013, 2013.

Woodhouse, C. A.: Winter climate and atmospheric circulation patterns in the Sonoran desert region, USA, Int. J. Climatol., 17(8), 859–873, doi:10.1002/(SICI)1097-0088(19970630)17:8<859::AID-JOC158>3.0.CO;2-S, 1997.

~~Table 1. Cross-correlations of Peak Month (PM) at GRDC stations for each classification technique for (a) observed and (b) simulated streamflow.~~

Classification Technique		5% Threshold	Q_{AMF}	Q_{7day}	Q_{15day}	Q_{30day}
Observed	5% Threshold	1				
	Q_{AMF}	0.866	1			
	Q_{7day}	0.894	0.912	1		
	Q_{15day}	0.895	0.880	0.945	1	
	Q_{30day}	0.900	0.832	0.881	0.890	1
Simulated	5% Threshold	1				
	Q_{AMF}	0.849	1			
	Q_{7day}	0.873	0.926	1		
	Q_{15day}	0.884	0.912	0.940	1	
	Q_{30day}	0.888	0.880	0.902	0.911	1

Table 1. Cross-correlations of Peak Month (PM) for each classification technique for observed and simulated streamflow where stations having different PMs.

Classification Technique		VBT1%	VBT3%	VBT5%	VBT10%	V01%	V03%	V05%	V10%	POT1	POT2	POT3
Observed	VBT1%	1.00										
	VBT3%	0.90	1.00									
	VBT5%	0.85	0.94	1.00								
	VBT10%	0.79	0.86	0.91	1.00							
	V01%	0.82	0.82	0.82	0.81	1.00						
	V03%	0.81	0.84	0.83	0.84	0.89	1.00					
	V05%	0.81	0.85	0.86	0.85	0.86	0.92	1.00				
	V10%	0.80	0.84	0.85	0.87	0.83	0.88	0.96	1.00			
	POT1	0.78	0.78	0.78	0.74	0.76	0.77	0.76	0.74	1.00		
	POT2	0.74	0.78	0.78	0.78	0.80	0.80	0.82	0.81	0.81	1.00	
	POT3	0.77	0.81	0.81	0.80	0.80	0.81	0.83	0.81	0.86	0.93	1.00
Simulated	VBT1%	1.00										

	VBT3%	0.87	1.00									
	VBT5%	0.83	0.95	1.00								
	VBT10%	0.80	0.88	0.90	1.00							
	V01%	0.86	0.85	0.84	0.84	1.00						
	V03%	0.87	0.86	0.85	0.83	0.92	1.00					
	V05%	0.87	0.88	0.85	0.84	0.90	0.97	1.00				
	V10%	0.82	0.87	0.86	0.85	0.83	0.89	0.92	1.00			
	POT1	0.80	0.83	0.83	0.81	0.83	0.86	0.86	0.82	1.00		
	POT2	0.78	0.81	0.80	0.79	0.79	0.83	0.83	0.82	0.92	1.00	
	POT3	0.80	0.81	0.79	0.80	0.80	0.83	0.84	0.81	0.92	0.95	1.00

Table 2. Average PAMF of each classification technique for modeled and observed where stations having different PMs.

Section	VBT1%	VBT3%	VBT5%	VBT10%	V01%	V03%	V05%	V10%	POT1	POT2	POT3
Observed	60.8%	61.7%	62.0%	62.0%	63.4%	63.6%	63.0%	62.5%	60.8%	59.1%	60.6%
Simulated	63.5%	64.5%	64.7%	63.5%	65.1%	64.8%	64.9%	64.1%	63.1%	60.3%	61.9%

Table 3. Percentage of stations according to the difference in PMs between modeled and observed streamflow at each classification technique.

Difference in PMs	VBT1%	VBT3%	VBT5%	VBT10%	V01%	V03%	V05%	V10%	POT1	POT2	POT3
Same	39%	39%	40%	42%	38%	39%	40%	42%	38%	36%	38%
≤±1 month	80%	81%	81%	80%	78%	79%	79%	79%	75%	75%	77%
≤±2 month	90%	91%	91%	90%	89%	90%	89%	89%	87%	87%	88%
≤±3 month	94%	95%	95%	95%	94%	95%	95%	95%	93%	93%	94%

~~Table 2.~~Table 4. Comparison of Peak Month (PM) for flooding and calculated P_{AMF} at 6 GRDC stations in the Zambezi River Basin.

Station (GRDC sta. numb.)	STA01 (1591001)		STA02 (1291100)		STA03 (1591406)		STA04 (1591404)		STA05 (1591403)		STA06 (1591401)		Final PM
Station name	Senanga		Katima Mulilo		Machiya Ferry		Kafue Hook Bridge		Itezhi-Tezhi		Kasaka		
River name	Zambezi		Zambezi		Kafue		Kafue		Kafue		Kafue		
Cumulative catchment area (km^2)	284,538		339,521		23,065		96,239		105,672		153,351		
Mean annual streamflow (m^3/s)	975		1168		139		287		353		988		
Streamflow type	Natural		Natural		Natural		Natural		Natural (Reservoir inflow)		Regulated		
Classification Technique	PM (month)	PAMF (%)	PM (month)	PAMF (%)	PM (month)	PAMF (%)	PM (month)	PAMF (%)	PM (month)	PAMF (%)	PM (month)	PAMF (%)	
Observed	4	96	4	100	3	93	3	100	3	94	7	36	3
Simulated	3	100	3	97	2	97	3	75	2	94	2	97	2

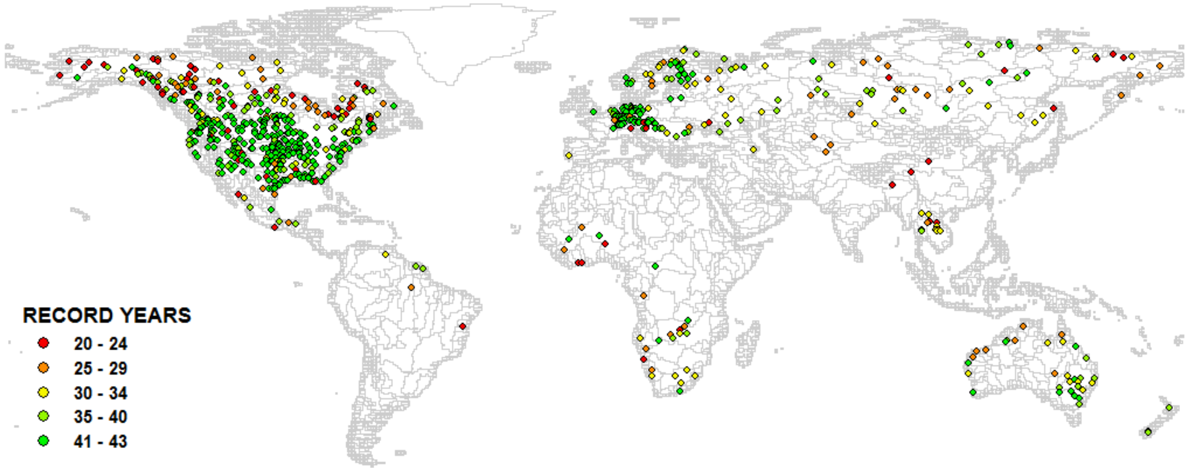


Figure 1. Location of 691 selected GRDC stations with corresponding number of years per station. Background polygons are world sub-basins based on 30′ drainage direction maps (Döll and Lehner, 2002) with separation of large basins (Ward et al., 2014).

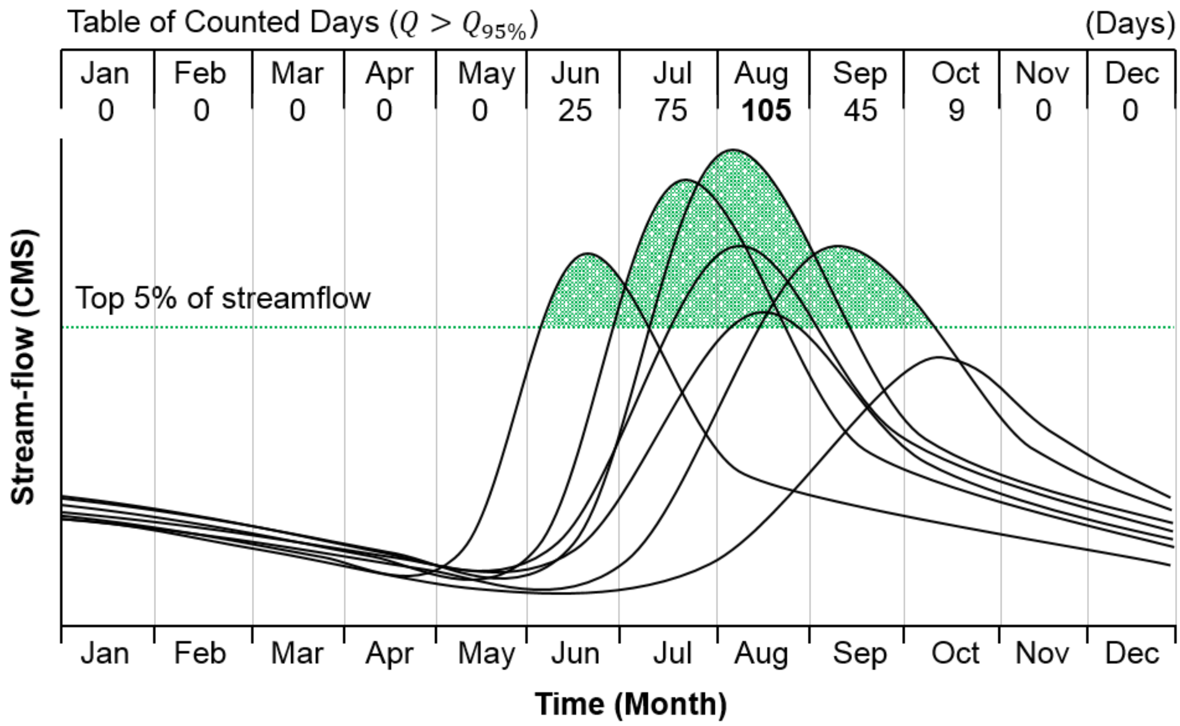


Figure 2. Seven years of synthetic streamflow data. Dotted line represents the 5% streamflow threshold. Numbers indicates the total days above the threshold for each month.

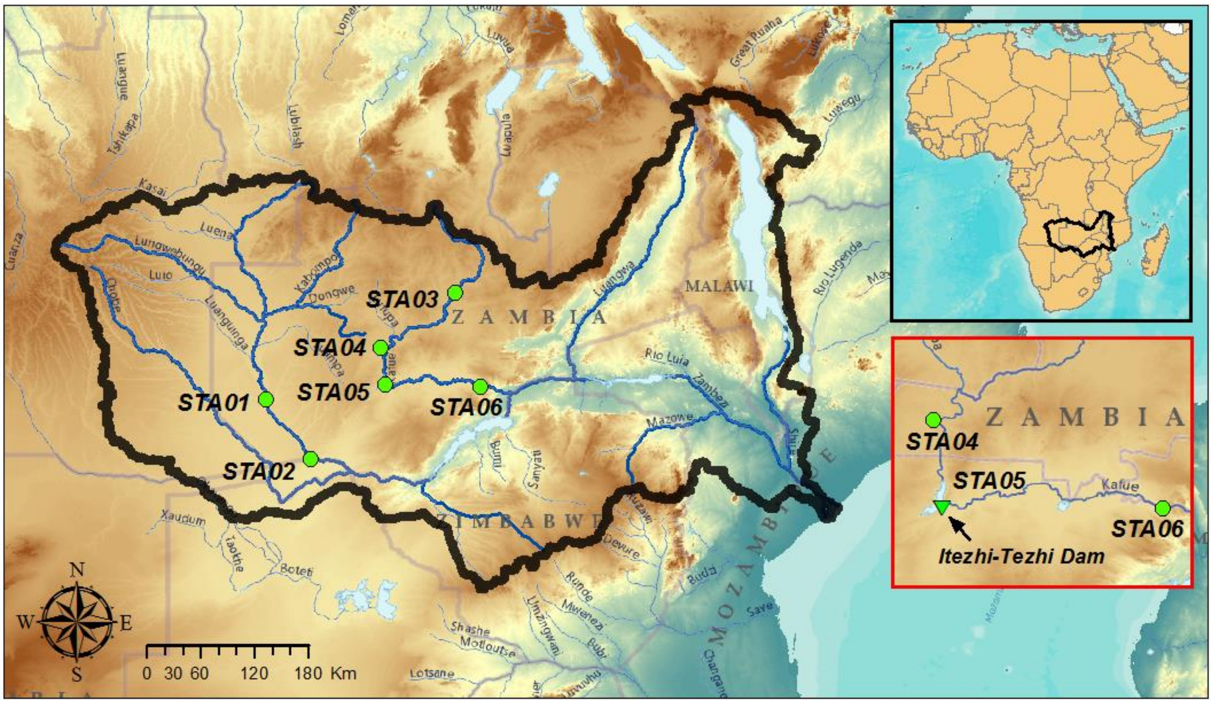


Figure 3. Map of Zambezi River Basin; the solid black line delineates the basin and the green points are the 6 GRDC stations (STA01-06), with STA06 downstream of the Itezhi-Tezhi dam (STA05.)

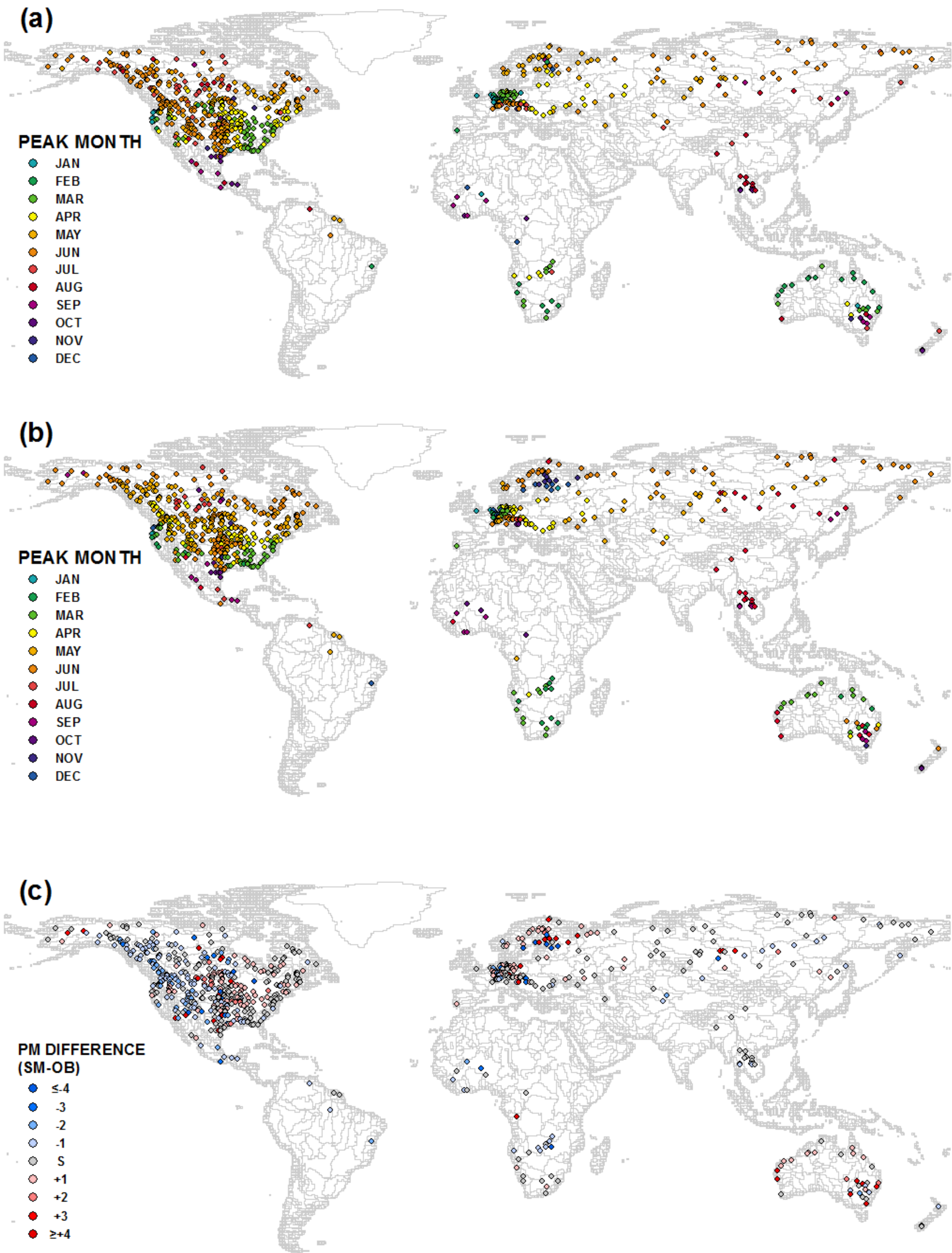


Figure 4. Peak Month (PM) for flooding as defined by (a) 691 GRDC observation stations, (b) simulated streamflow at associated locations and (c) Temporal difference in PM between observations and simulation (SM-OB, number of months).

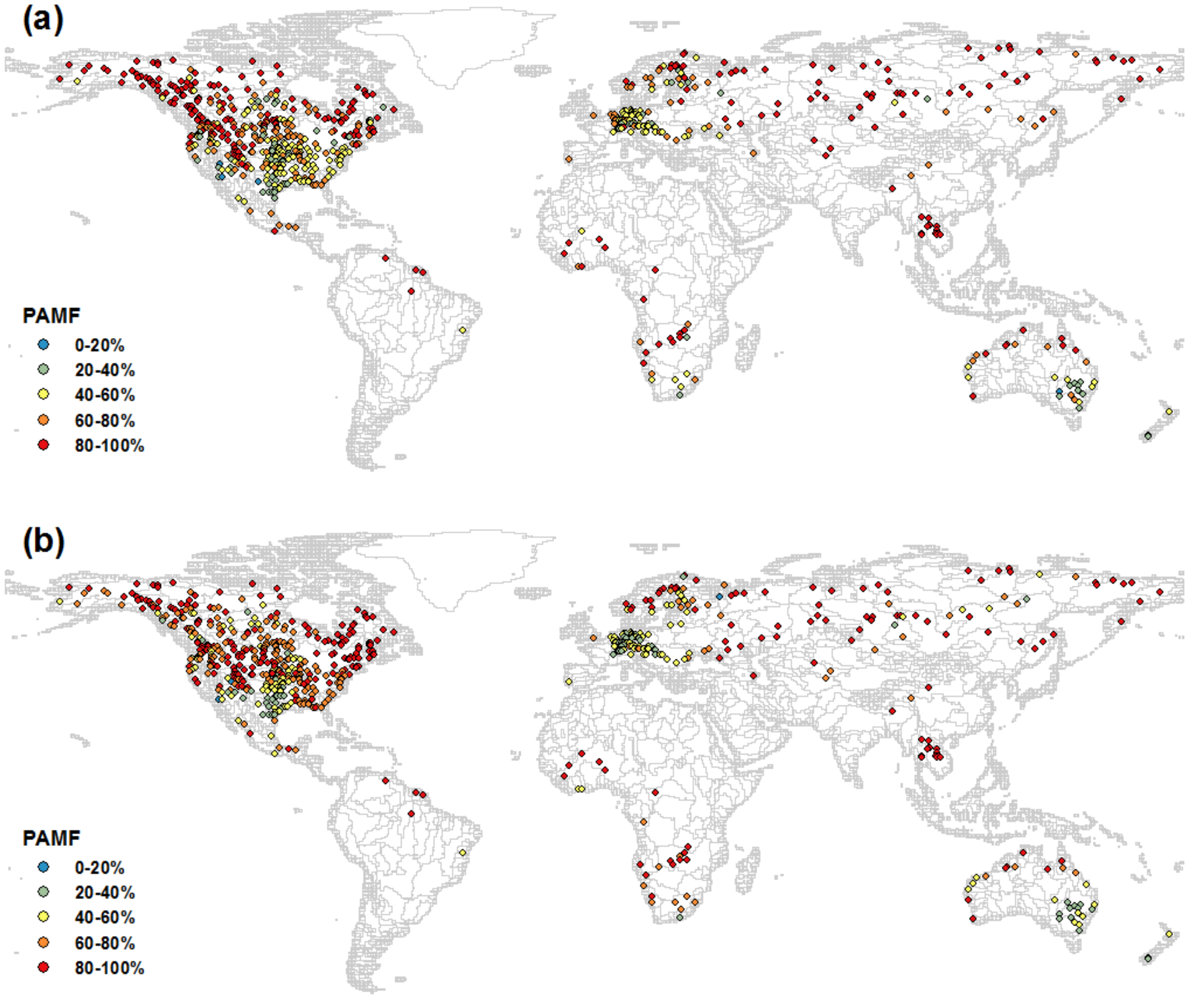


Figure 5. Calculated Percentage of Annual Maximum Flow (PAMF) values for (a) 691 GRDC observation stations, and (b) simulated streamflow at associated locations.

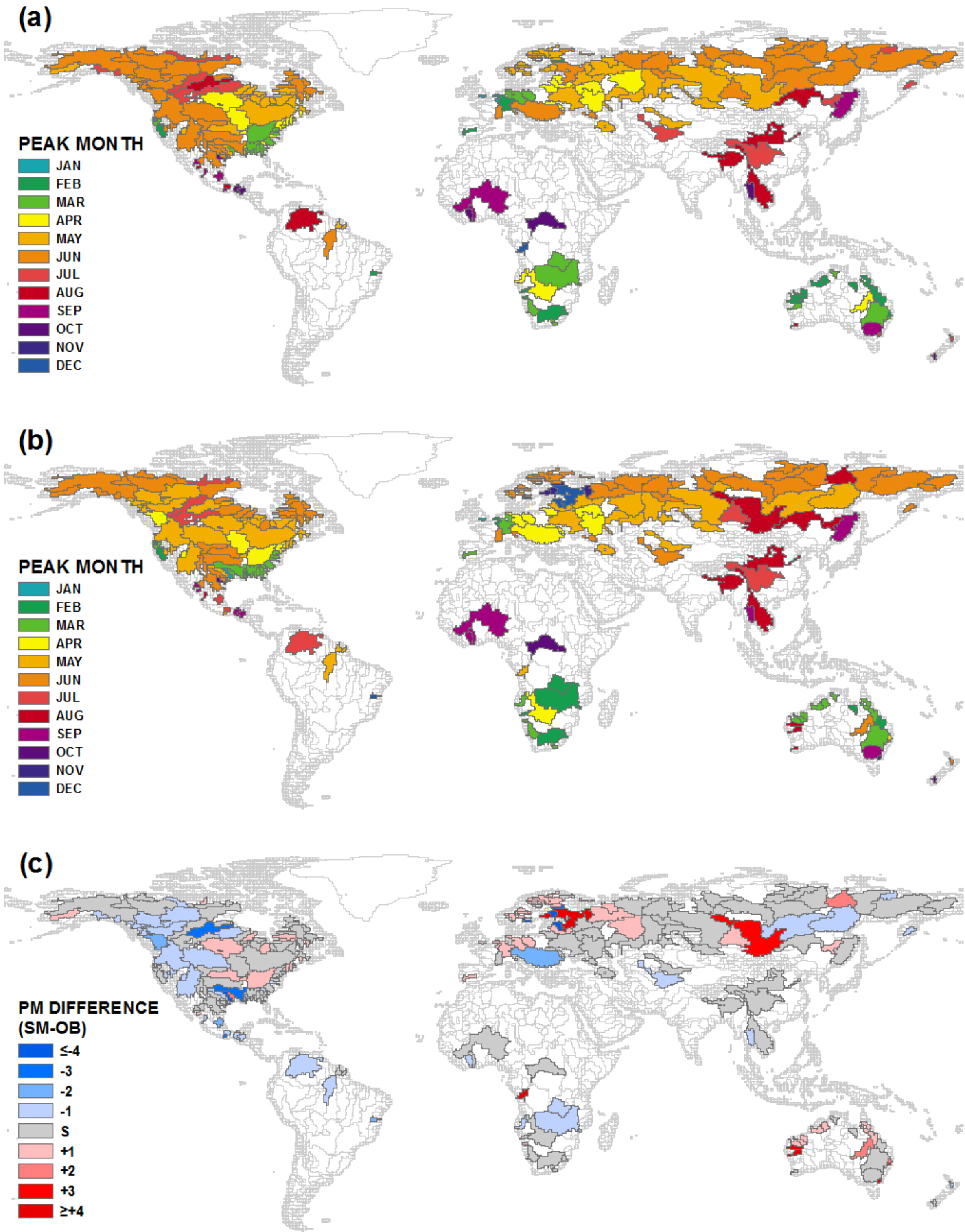


Figure 6. Peak Month (PM) for flooding by sub-basin as defined by (a) 691 GRDC observation stations, (b) simulated streamflow at associated sub-basins and (c) Temporal difference in PM between observations and simulation (SM-OB, number of months).

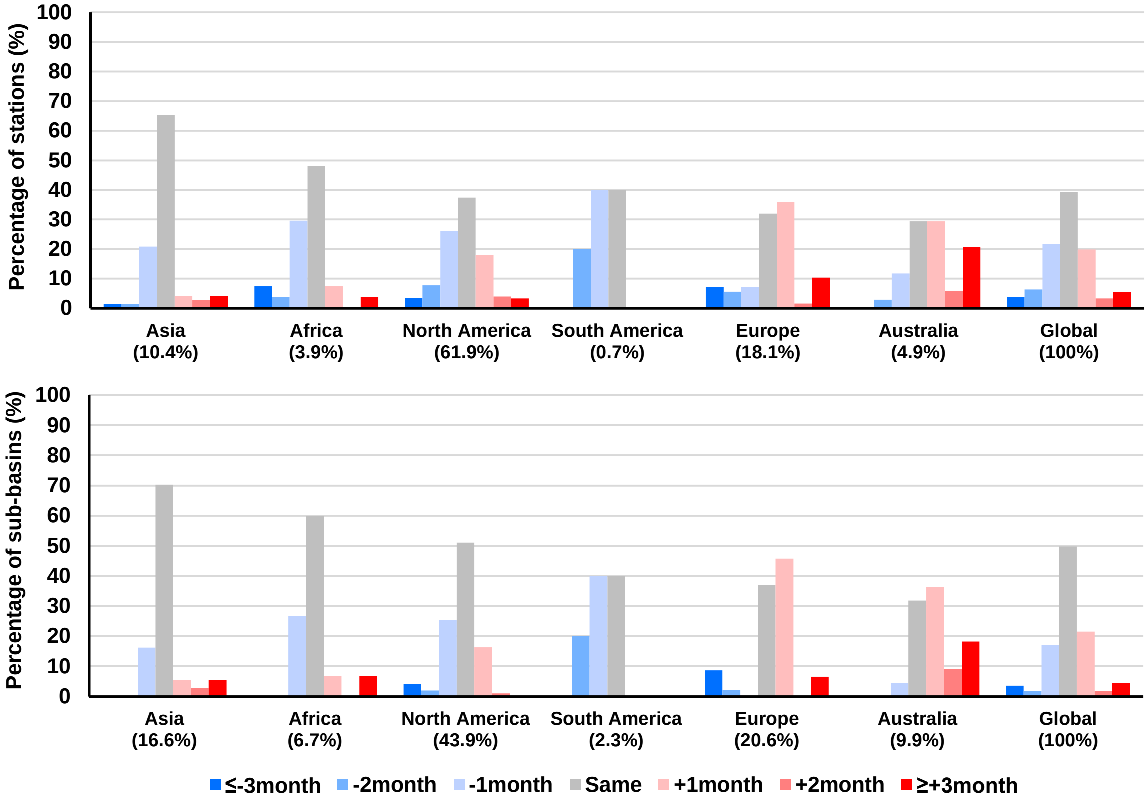
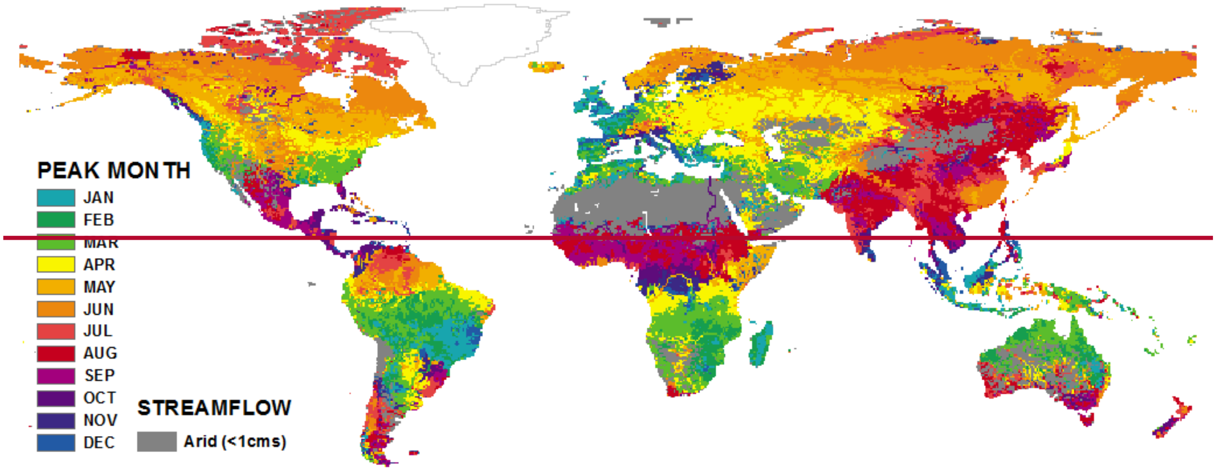
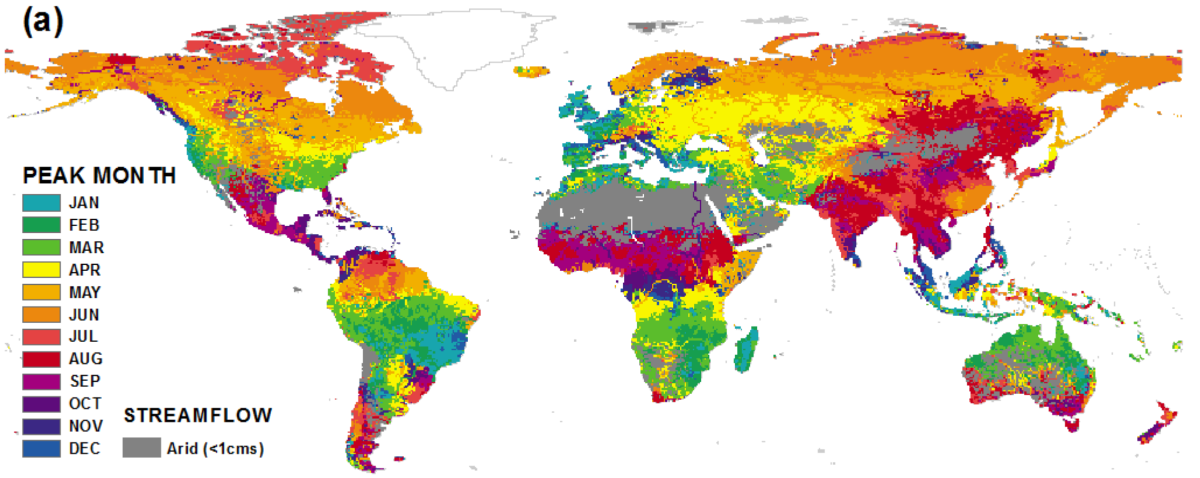
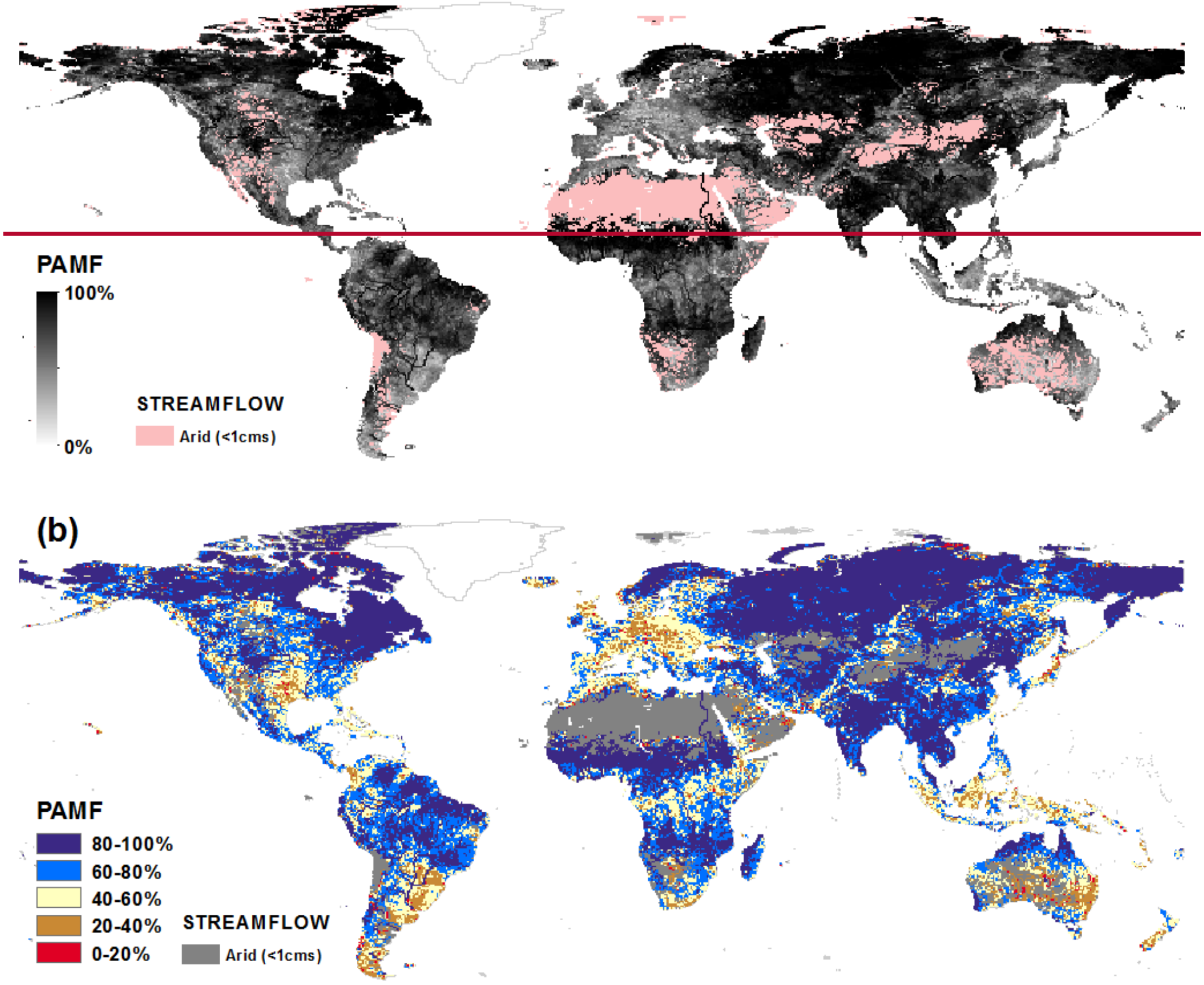


Figure 7. Percentage of stations (above) and sub-basins (below) according to temporal difference of PM between observations and model outputs (SM-OB, number of months) in each continent.


PEAK MONTH
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
STREAMFLOW
Arid (<1cms)


(a)
PEAK MONTH
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
STREAMFLOW
Arid (<1cms)



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

~~Figure 9. Calculated Percentage of Annual Maximum Flow (PAMF) values for at all modeled grid cells.~~ Figure 8. (a) Peak Month (PM) as defined at all modeled grid cells (b) Calculated Percentage of Annual Maximum Flow (PAMF) values for at all modeled grid cells; subjectively classified as high = 80-100%, moderate = 60-80%, low = 40-60% and poor = 0-40%.

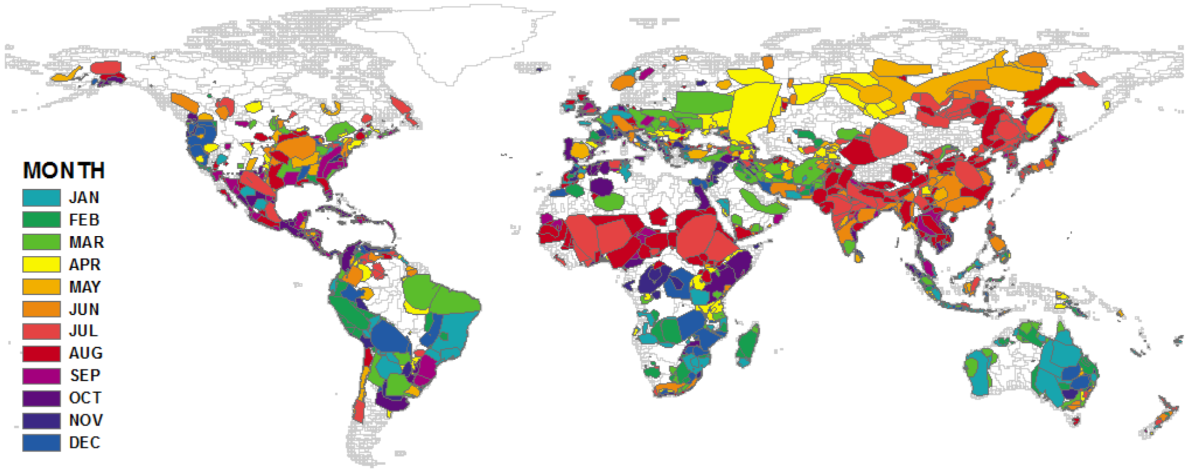


Figure 9~~10~~. Archive of major flood events globally from the Dartmouth Flood Observatory (DFO) over 1985-2008.

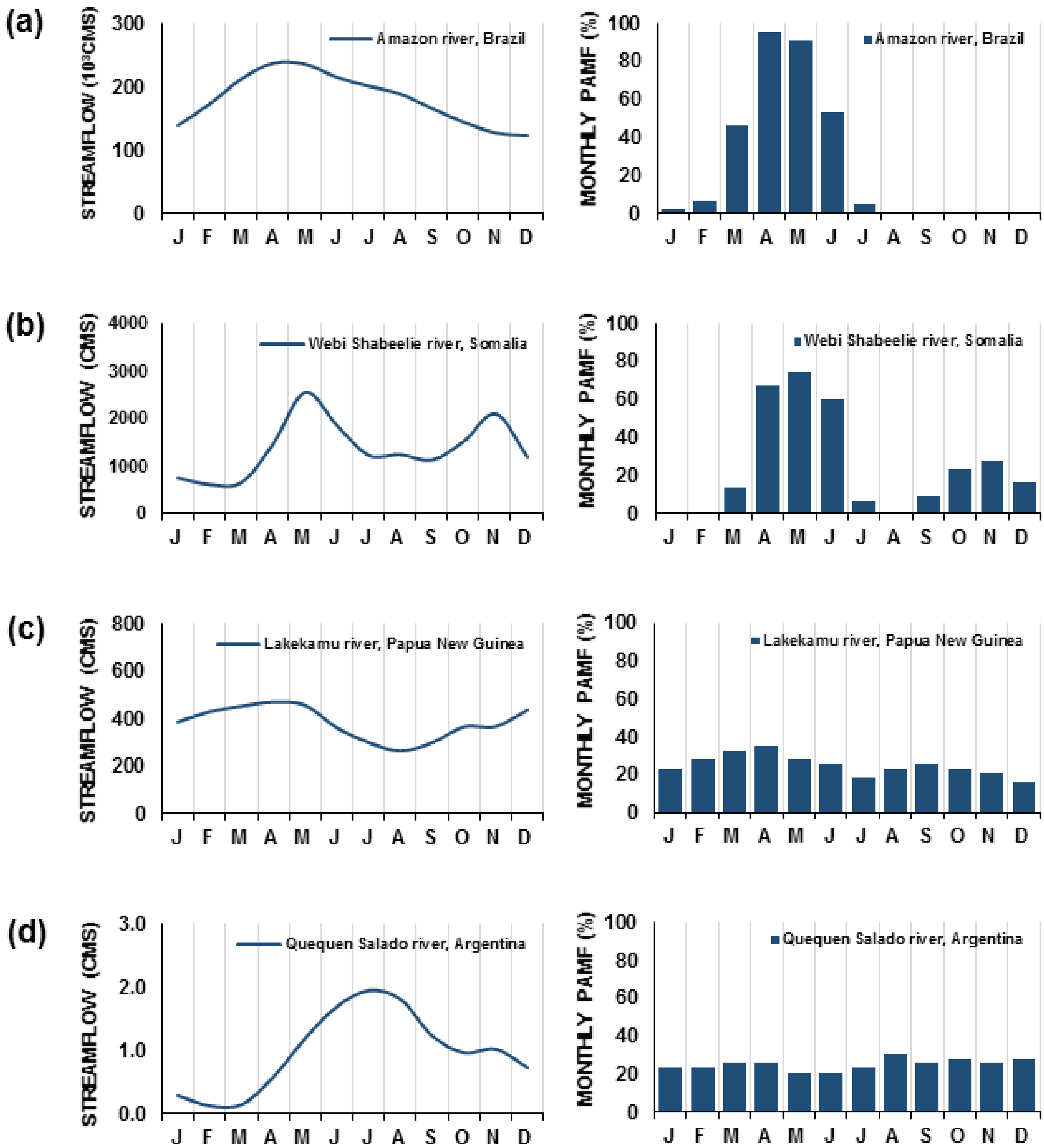


Figure 10~~11~~. Model-based streamflow climatology (left) and corresponding monthly PAMF (right.) Types and locations are: a) uni-modal streamflow – At Bom Lugar, Amazon river, Brazil, b) bimodal streamflow – At Saacow, Webi Shabeelie river, Somalia, c) constant streamflow – At Terapo Mission, Lakekamu river, Papua New Guinea and d) low-flow – At La Sortija, Quequen Salado river, Argentina.

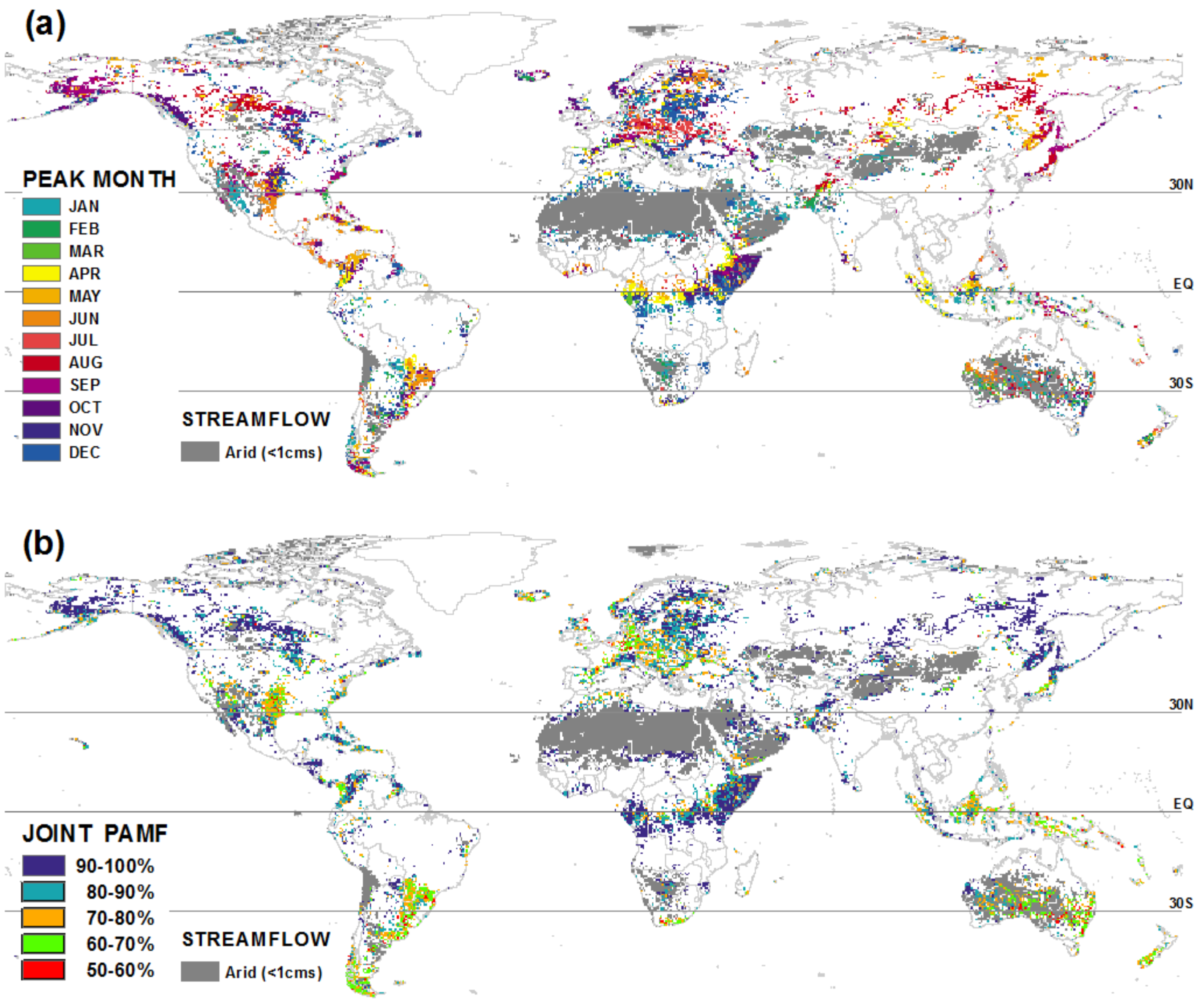


Figure 1112. (a) Minor Peak Month (PM) for flooding as defined at detected grid cells and (b) joint PAMFs of major and minor PMs at corresponding cells.

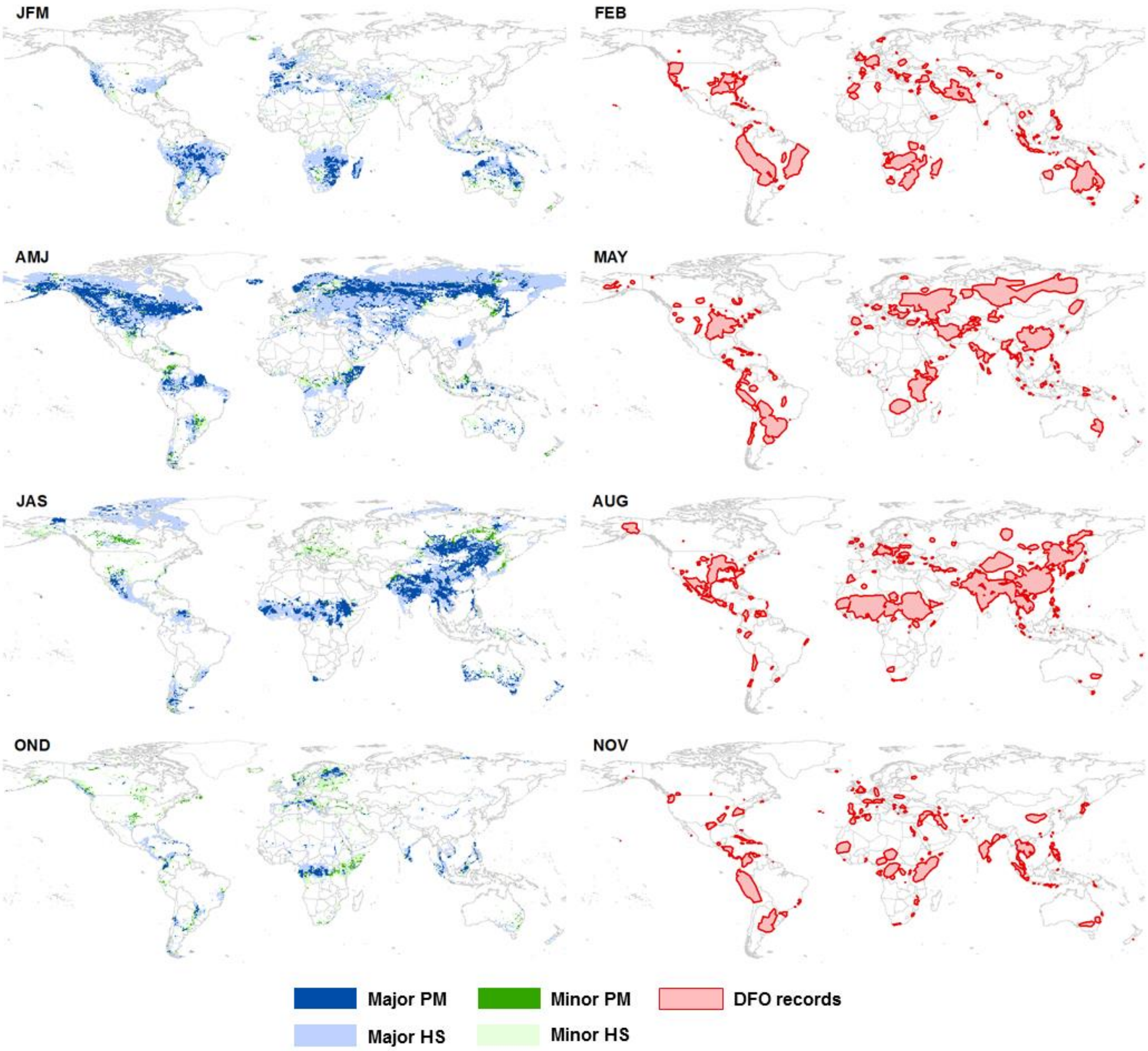


Figure 12~~13~~. Defined major ~~FS~~HS and minor ~~FS~~HS where joint PAMF is greater than 60% (left); peak month of major and minor ~~FSs~~HSs (dense color) and pre- and post-month of major and minor ~~FSs~~HSs (light color.) Monthly accumulated actual flood records (DFO) during 1958-2008 (right.)