Global-Scale Prediction of Flood Timing Using Atmospheric Reanalysis

Hong Xuan Do[1,3,4], Seth Westra[1], Michael Leonard[1], Lukas Gudmundsson[2]

[1]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, Australia

[2] Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, Universitaetstrasse 16, Zurich 8092, Switzerland.

[3]Faculty of Environment and Natural Resources, Nong Lam University, Vietnam

[4]School for Environment and Sustainability, University of Michigan, Ann Arbor, United States.

Corresponding author: Hong Xuan Do (hong.do@adelaide.edu.au)

Key Points:

- Observation-based analysis of the timing of annual maximum streamflow indicates large-scale patterns over the common 1981-2010 period
- Regional patterns of flood generation mechanisms are highlighted through a comparison between high-flow timing and the timing of seven climate predictors
- A prediction of flood timing was made for the global land mass using an atmospheric reanalysis dataset

Abstract

The annual timing of flood events is a useful indicator to study the interaction between atmospheric and catchment processes in generating floods. This paper presents an assessment of the seasonal timing of floods for 7,894 gauging locations across the globe over a common period from 1981 to 2010. The averaged ordinal date of annual maximum streamflow is then estimated for ungauged locations following a two-stage prediction scheme. The first stage identifies regions that share a common climatic predictor of flood timing by analysing the similarity of flood timing with seven climate variables. These variables represent precipitation timing and snow melt dynamics, and are derived from a global climate reanalysis dataset. Homogeneous regions in terms of the dominant predictor are generalised in the second stage through a rule-based classification. The classification partitions the world into ten hydro-climate classes, where each class has flood timing predicted using the most relevant climate predictor. Using this relatively simple and interpretable model structure, flood timing could be predicted with a global mean absolute error of approximately 32 days, whilst maintaining consistency across large regions. Potential applications of the developed map include better understanding of climatic drivers of flooding, and benchmarking the performance of global hydrological models in simulating the processes relevant to flooding.

Plain Language Summary

Timing of annual maximum streamflow is a useful index to relate flood occurrence to appropriate flood generation processes. This study presents an assessment of flood timing across 7,894 gauging locations globally over the period from 1981 to 2010. The averaged date of annual maximum streamflow is compared to seven climate predictors, identifying regions that are likely to share a common flood generation process. These homogeneous regions are generalised across the globe using a gridded dataset of daily precipitation and temperature. To derive a global map of flood timing, the date of annual maximum streamflow is predicted for both gauged and ungauged locations, using a linear function of the most important climate predictor in each region.

1 Introduction

The seasonal timing of flood events is a useful indicator of how atmospheric processes interact with the local catchment, with recent papers showing the relevance of intense precipitation, snow melt and rain-on-snow events as mechanisms driving the timing of floods [*Blöschl et al.*, 2017; *Hall and Blöschl*, 2018; *Iliopoulou et al.*, 2019; *Parajka et al.*, 2010; *Villarini*, 2016]. An understanding of flood timing provides useful insights at many scales: (i) globally – because of the considerable attention devoted to the development of global hydrological models [*Bierkens*, 2015; *Bierkens et al.*, 2015; *Wood et al.*, 2011], and the need to reconcile patterns of non-stationarity in climatic drivers such as rainfall [*Sharma et al.*, 2018; *Westra et al.*, 2013; *Westra et al.*, 2014] with those observed in streamflow [*Do et al.*, 2017; *Gudmundsson et al.*, 2017; *Gudmundsson et al.*, 2019; *Hodgkins et al.*, 2017]; (ii) regionally – for analyses of flood frequency within homogeneous regions and for detection/attribution of historical changes in flooding [*Cunderlik et al.*, 2004; *Villarini*, 2016]; and (iii) locally – to assist understanding of flood mechanisms, as required by decision makers in designing strategies for flood prevention, mitigation, and response [*Dhakal et al.*, 2015; *Ward et al.*, 2015].

There have been many studies of flood magnitude and frequency characteristics at global [*Asadieh et al.*, 2016; *Dankers et al.*, 2014; *Do et al.*, 2017; *Do et al.*, 2019; *Hodgkins et al.*, 2017; *Wasko and Sharma*, 2017; *Woldemeskel and Sharma*, 2016], continental [*Alfieri et al.*, 2015; *Gudmundsson et al.*, 2012; *Hall et al.*, 2014; *Ishak et al.*, 2013; *Mallakpour and Villarini*, 2015; *Mediero et al.*, 2015; *Parajka et al.*, 2010] and national scales [*Beurton and Thieken*, 2009; *Burn and Whitfield*, 2016; *Merz et al.*, 2018; *Slater and Villarini*, 2016; *Stevens et al.*, 2016]. However, there have been comparatively fewer and more recent studies of flood timing [*Berghuijs et al.*, 2016; *Berghuijs et al.*, 2019; *Blöschl et al.*, 2017; *Burn and Whitfield*, 2016; *Cunderlik and Ouarda*, 2009; *Dettinger and Diaz*, 2000; *Hall and Blöschl*, 2018; *Villarini*, 2016; *Ye et al.*, 2017]. Interestingly, most of the studies of flood timing find unique information on the mechanisms that cause floods. In particular, unlike indicators of flood magnitude or frequency, the average timing of floods is relatively independent of human influences including land use change and river regulation [*Villarini*, 2016]. A regional investigation [*Hall and Blöschl*, 2018] also found that geographical location is potentially the dominant factor driving flood seasonality.

Most studies of flood timing have focused on Europe and North America, so that a global perspective of when and why floods occur at different times of the year is not yet available. To develop this global perspective, it is essential to expand the assessment of flood timing to other continents (e.g. Australia, South America, Asia, Africa) using a consistent dataset and analysis methodology. One possibility is to simulate runoff and extract information of flood timing through the use of global hydrological models [*Lee et al.*, 2015] forced with global reanalysis climate. To our knowledge, *Lee et al.* [2015] is the only model-based study to produce a global map of the peak flow season (defined as the consecutive three-month period with the highest number of events above a threshold of streamflow volume), whereas model-based studies of timing of annual maximum streamflow are not yet available. More recently, *Ghiggi et al.* [2019] provides global maps of the month with minimum and maximum flow, based on a data-driven century long runoff reconstruction. Another alternative possibility is to estimate flood timing using available observational datasets from across the globe, followed by the construction of a data-driven model to infer flood timing at locations without streamflow observations. In addition to providing meaningful information in its own right, such an approach would provide a useful point of comparison for any subsequent model-derived maps of flood timing.

The spatial variation of the dominant mechanisms in flood generation, however, poses a challenge to predicting flood timing for ungauged locations. Heavy rainfall is one of the most common sources of flooding, as the catchment rapidly saturates due to receiving a significant amount of precipitation [*Kozlowski*, 1984]. Many studies have shown that other factors also play an important role in the flood generation processes, including antecedent soil moisture [*Bennett et al.*, 2018; *Ivancic and Shaw*, 2015; *Wasko and Sharma*, 2017; *Wasko and Nathan*, 2019; *Ye et al.*, 2017] and snowmelt dynamics [*Berghuijs et al.*, 2016; *Blöschl et al.*, 2017; *Mediero et al.*, 2015; *Parajka et al.*, 2010]. Flooding in arid regions or very large catchments may be more sensitive to the total amount of rainfall over long periods (up to months) rather than short duration rainfall events [*Ingle Smith*, 1999; *Johnson et al.*, 2016; *Marengo*, 2006; *Pathiraja et al.*, 2012], and thus the long-term total precipitation also needs to be taken into account. A reliable model for flood timing, therefore, must possess the capacity to identify regions in terms of the dominant flood generation process(es), which can then be used to predict flood timing in ungauged locations.

The recent publication of a global archive of over 30,000 streamflow gauges [GSIM; *Do et al.*, 2018b; *Gudmundsson et al.*, 2018b] provides a unique opportunity to explore many aspects of streamflow characteristics at the global scale, including flood timing. The main aim of this study is to use this resource, combined with an atmospheric reanalysis dataset, to develop a data-driven model to infer flood timing at both gauged and ungauged regions across the globe. Specifically, the global seasonality of flood timing is first evaluated across all GSIM stations with sufficient data. Observations of flood timing are then analysed with respect to seven climate predictors derived from the reanalysis dataset to identify potential flood producing mechanisms. The single predictor best suited to explain and predict flood occurrence is then identified at each location.

The regional consistency between flood timing and the most relevant predictor is then generalised to all gauged and ungauged locations using a rule-based classification system that identifies homogenous regions in terms of the predictor-flood timing relationships.

The remainder of this paper is structured as follows. Section 2 provides an overview of the data and methods that were used to assess the seasonal timing of floods and the prediction scheme development. The results are reported in Section 3 together with discussions about the performance of the prediction scheme. Finally, Section 4 summarises the key findings and highlights potential application of the proposed prediction scheme.

2 Data and methods

This section summarizes the workflow for global prediction of flood timing (Figure 1), including input variables, observational analyses, and the prediction scheme using a rule-based classification system. The datasets used in this study are presented in section 2.1, followed by descriptions about observations and predictors of flood timing (section 2.2) and the development of a prediction scheme to derive a global map of flood timing (section 2.3).

2.1 Datasets

The Global Streamflow Indices and Metadata (GSIM) archive contains streamflow indices from more than 30,000 stations across the globe [*Do et al.*, 2018b; *Gudmundsson et al.*, 2018b]. To establish a compromise between data quality and availability, only stations classified with a ‘useful’ homogeneity class [*Gudmundsson et al.*, 2018b] are used to ensure that stations with potentially spurious step changes are excluded. A threshold of at least 20 yearly data points available during the 1981-2010 common period (with each year having at least 350 days of reliable records) was used to select streamflow gauges with sufficient data to minimise the influence of inter-annual and inter-decadal variability, while maintaining a relatively large sample for a global scale investigation. This filtering process identified 9,560 viable stations, of which a further 76 stations were removed due to unavailability of catchment area information.

To mitigate the influence of large-scale climate gradients as well as routing effects and catchment processes, an approach of previous global-scale reconstruction studies was adopted [*Beck et al.*, 2015; *van Dijk et al.*, 2013], and only stations with catchment area less than 10,000km^2 were retained—approximately the size of a one-degree longitude/latitude grid cell This led to the removal of a further 1,226 stations. Finally, 364 stations that fall outside of ERA-Interim land regions were also removed (primarily over coastal regions or islands), as it was not possible to develop predictions for these locations. The outcome of this filtering process was the identification of a final subset of 7,894 stations (out of the original 30,959 GSIM stations) to be used for this study.

To represent global observations of atmospheric forcing, the ERA-Interim dataset was used over the same 1981-2010 period [*Dee et al.*, 2011]. Regridded daily temperature and precipitation data products at 0.5-degree resolution were retrieved directly from the European Centre for Medium-Range Weather Forecasts data portal. The land-sea mask from ERA-Interim was used to keep only values over land regions (except Greenland and Antarctica, which were excluded). Time series at monthly and annual resolutions were aggregated from original daily time series. The empirical analysis was conducted at each streamflow gauge, and thus reported streamflow gauge coordinates [*Do et al.*, 2018a] were used to identify corresponding grid-cells from the global climate dataset and extract information of both precipitation and temperature for each streamflow station. Note that although the streamflow data has inconsistent coverage across the globe, the reanalysis data covers all the global land areas, and thus provides the capacity to extrapolate the findings at data-covered regions, and make predictions of flood timing across the globe.

2.2 Observations and predictors of flood timing

2.2.1 Observations of flood timing

The ordinal day (from 1 to 365 or 366, and starting on 1 January) of annual maximum streamflow (DOYMAX index, available in GSIM [*Gudmundsson et al.*, 2018a]) was selected as the indicator of flood seasonality. Circular statistics [*Blöschl et al.*, 2017; *Mardia and Jupp*, 2009] were used to assess the seasonality of historical flood timing, with further details provided in the supplementary material. As the present study focused on assessing long-term-mean of flood timing, the circular-mean value of each DOYMAX time series over the period from 1981-2010 was used as the observed timing of flood seasonality for each stream gauge. A concentration statistic (R) of each DOYMAX time series was also calculated to represent the strength of the seasonality, where $R = 0$ indicates that flood occurrence dates were spread evenly throughout the year, and $R = 1$ indicates that all flooding events occur on the same ordinal day across all years.

Note that a low value of the flood timing concentration index R does not always correspond to low levels of seasonality and could reflect other complex flood timing distributions (e.g. reflective symmetric bi-modal, or asymmetric unimodal), which is beyond the scope of our investigation. Stations with non-seasonal flood timing were identified through a circular Kuiper's test, which evaluates whether the time series is circularly uniform. Only stations for which the null hypothesis of circular uniformity is rejected at the 10% confidence level (i.e. those stations that have statistically significant seasonality) were considered as input for the prediction of flood timing (7,040 stations in total).

Figure 2 shows examples of calculating the mean and concentration of flood timing. Figure 2a illustrates a location where flood events can occur at any time of the year. The hypothesis of uniformity was not rejected at the 10% significance level in this case, suggesting the absence of flood seasonality evidence. Figure 2b provides an example of seasonality where all flood events occur between November to April and the majority of the events fall in January and February.

2.2.2 Predictors of flood timing

This section presents seven identified climate predictors of flood timing considered in this analysis. To ensure global availability for the prediction, daily precipitation and temperature data at each grid point of the ERA-Interim datasets were used to derive the identified predictors. Each predictor is the circular mean value of the occurrence date of one hypothesised flood generation process over the 1981-2010 period. The seasonality assessment using the circular uniformity hypothesis was also applied to these seven predictors, so that only grid cells where the null hypothesis of circular uniformity was rejected at the 10% level are considered, while a missing value was assigned for other grid cells. The seven climate predictors are divided into three groups based on the hypothesised flood generation processes that they represent.

The first group of predictors focus on short-term rainfall and reflect the hypothesis that heavy rainfall events are the primary mechanism driving large streamflow events. Based on the contributing areas of the gauging stations, it is estimated that all stations in the final subset (which by design have catchment areas less than 10,000km^2) had times of concentration of seven days or less – based on the Pilgrim McDermott formula [*Pilgrim et al.*, 1987]. This suggests that heavy rainfall events spanning a period of seven days or less are appropriate to represent this mechanism. Two variables were extracted from ERA-Interim precipitation dataset considered in this group:

(i) PD - date of peak daily precipitation in each calendar year, which represents flood produced by the single largest rainfall event, and

(ii) PD7 - date of peak 7-day precipitation in each calendar year, which represents flood produced by the largest series of rainfall events. To extract PD7, a backward-moving window of seven days was applied to each day of the year, and the day of maximum value (comprising the total precipitation depth on that day and the six prior days) was recorded for each calendar year.

The second group of predictors focus on long-term rainfall and reflect the hypothesis that long-term catchment wetness and antecedent moisture conditions play a key role in the flood generation process. There are two variables considered in this group:

(i) PD30 - the date of peak 30-day precipitation in each calendar year, which represents the hypothesis that the peak discharge occurs when the drainage area is relatively wet, and

(ii) PD90 - the date of peak 90-day precipitation in each calendar year, which represents the hypothesis that the timing of peak discharge occurs toward the end of a wet season, where significant build-up of catchment moisture will have occurred. The calculation process for PD30 and PD90 was similar to PD7, but with the backward-moving window set at 30 days and 90 days respectively.

The third and final group of predictors focus on snowmelt processes and are designed to provide an indicator of snowmelt or rain-on-snow processes. The first predictor in this group is the date of seasonal transition from snowfall to rainfall in precipitation (TD), which is defined as the first seven-day (the first day was chosen as TD value) that have averaged surface air temperature rises above $0^{o}C$ after having been below $0^{o}C$ for at least seven consecutive days. To represent a more sophisticated indicator for snowmelt events, a simple degree-day method [*Berghuijs et al.*, 2016; *Hock*, 2003; *Woods*, 2009] was used to simulate snow-dynamics (see supplementary material for detail methodology). This led to two predictors derived from simulated snowmelt contribution:

(i) SD – the date of peak value of daily snowmelt or rain-on-snow

(ii) SD7 – the date of the peak value of 7-day snowmelt or rain-on-snow. Here a backward-moving window of seven days was used to calculate the time series of total snowmelt or rain-on-snow amount.

To mask out locations where there was an absence of significant contribution of snowmelt to flood generation, additional constraints were applied to snowmelt predictors. For the TD predictor, locations where this variable cannot be identified for more than 70% of the years were assigned a missing value. Missing values were also assigned to SD and SD7 predictors across locations where less than 10% of precipitation falls as snow.

The availability of chosen climate predictor groups across the globe is shown in Figure 3. The constraining criteria for snowmelt predictors imply that these predictors are mostly available in high latitude regions in the northern hemisphere, and in some mountainous areas in the southern hemisphere such as the Andes in South America and the Southern Alps in New Zealand. Due to non-seasonality of the selected predictors (i.e. where the circular Kuiper test does not reject the uniformity hypothesis), some areas, mostly desert regions, do not have any available predictors, such as in the interior of southern Australia, the south-eastern part of the Arabian Peninsula, or the Uruguay River. Furthermore, no rainfall predictors are available for many grid cells across the Appalachian Mountains in North America, Eastern Europe, central Kazakhstan, and northern Africa as a result of the lack of rainfall seasonality in these regions (i.e. the null hypothesis of circular uniformity is not rejected at the 10% confidence level). Detailed maps of the timing and seasonality of each predictor are provided in the supplementary materials (Figures P1 to P7).

2.3 Developing a global prediction of flood timing

At the global scale, it has been shown previously that the mechanism dominating flood occurrence varies significantly in many regions [*Berghuijs et al.*, 2016; *Blöschl et al.*, 2017], and thus a reliable prediction for flood timing must adequately reflect this spatial variation. To facilitate this requirement, the present study proposes a two-stage prediction model, in which the first stage (sub-section 2.3.1) aims to define homogeneous regions in terms of the most important predictor. In the second stage (sub-section 2.3.2), the defined homogeneous regions are generalised across the globe through a classification scheme, in which prediction of flood timing is made for each class by a linear function of the most relevant predictor. The global prediction for flood timing was then obtained by applying the classification system and the linear functions to all land locations, including ungauged regions.

2.3.1 Diagnostic of regional consistency between predictors and observed flood timing

This section describes the first stage of the flood timing prediction scheme, aiming to define regional patterns of the most important flood timing predictors from observational data. The temporal discrepancy between the average ordinal dates of predictors and annual maximum streamflow events was first calculated to identify the climate variable with the closest match at each location across selected stations showing seasonal flood timing. The circular characteristic of the variables was also considered, allowing for the discrepancy between 31st December and 1st January to be one day rather than 364 days (see supplementary material for calculation of circular statistics). The level of consistency between flood timing and the predictor with the closest match was assessed by grouping stations into five categories based on the magnitude of temporal discrepancy, as outlined in Table 1. The spatial distribution of the single predictor with the highest level of consistency to flood timing at each gauged location was then used to represent homogeneous regions in terms of the dominant climate predictor.

2.3.2 Predicting flood timing using a rule-based hydro-climate classification

In the second stage of the prediction scheme, the observed homogeneous regions were generalised across the globe through a rule-based classification system, which used a set of climate indices (derived from ERA dataset and represent the climatic conditions across the world) as separating variables. The indices were derived by first using nine variables from the Köppen-Geiger [*Köppen*, 1900] classification, which is currently the most widely used climate classification system. In addition, we also calculated the fraction of total precipitation that falls as snow (f_{snow}) and an indicator of whether transition time from snowfall to rainfall can be reliably identified ($TD_{indicator}$) to support the development of a classification tree (i.e. to better delineate the boundary between snowmelt dominant and rainfall dominant flood regions). This led to the selection of 11 'candidate' indices, which are summarised in Table 2.

The classification scheme has a similar structure to that of a classification tree, which is a binary tree with nodes defined by simple splitting rules applied to a set of input variables and corresponding thresholds (e.g. at the root node, all stations are divided into two groups by a decision rule 'transition time from snowfall to rainfall can be reliably identified'). However, each leaf of the tree (i.e. terminal node or hydro-climate class in the context of this study) provides a prediction of flood timing through a linear function of one of the seven climate predictors, rather than the output of the tree simply being the assignment of a class.

To develop this classification scheme, one possible option is to apply machine learning techniques such as recursive binary splitting together with a greedy pruning algorithm (see *Cannon* [2012] for example) on available datasets. However, we decided to construct the model in a semi-automated manner to ensure that the final hydro-climate classification system is as physically interpretable as possible, while retaining regional patterns of predictors that best explain the occurrence of flood.

Figure 4 illustrates our approach, with two different procedures were applied for 'non-terminal' and 'terminal' nodes separately. At each non-terminal node *n*, a specific climate index (selected from 11 climate variables and denoted here by C_n) and corresponding thresholds were manually selected to divide the world into sub-regions. To guide the selection of climate index C_n, visual matching was first conducted between the spatial variations of all climate indices (see supplementary Figures C1 to C11) to the regional consistency between predictors and observed flood timing (results of the method presented in Section 2.3.1; an example is provided in the methodology section of the supplementary). This step identified the climate indices that can potentially serve as splitting variables to divide the world into hydro-climate classes, with each class sharing a common flood predictor. Among the "short-listed" climate indices, the variable and associated threshold that could be meaningfully linked to a flood generation mechanism (e.g. snowmelt processes, heavy rainfall events and long-term catchment wetness) were then chosen.

We define the terminal node of the partitioning scheme (hydro-climate class R_j, where j is the index over all classes) to be a homogenous 'region' that shares a common flood timing predictor. At each terminal node R_j, the timing of the flood (denoted by Y_{R_j}) is then predicted by adding a lag-day (denoted by γ_{R_j}) to the value of the best climate predictor (denoted by X_{R_j}, which is one of the seven climate predictors (i.e. indicators of rainfall and snow melt timing) that are defined in Section 2.2.2). For a specific hydro-climate class R_j, the prediction of flood timing ($\hat{Y}_{R_j}$) was made using a linear equation:

$$\hat{Y}_{R_j} = X_{R_j} + \gamma_{R_j} \qquad j = \{1,2,\dots,J\} \qquad (1)$$

The central idea of this prediction scheme is that regions with the same hydro-climate class (R_j) are likely to have floods, on average, occurring γ_{R_j} days within the occurrence of the hypothesised mechanism (γ_{R_j} is bounded between -15 and +15 days to prioritise predictor that has a high consistency to flood timing). For example, if the peak daily precipitation (PD) was identified as the most suitable predictor for hydro-climate class R_j, the flood timing at each station in this class is predicted by adding a constant γ_{R_j} to the date of the peak daily precipitation.

The best predictor and corresponding lag-day for each hydro-climate class was determined following an automated optimisation. The objective of the optimisation was to (i) minimise error between predicted and observed flood timing and (ii) maximise the proportion of locations that have available data for the predictor. The former criterion represents the predictive ability while the latter indicates the ability to correctly identify regions with common flood generation processes of the prediction scheme. The objective function used an adjusted mean absolute error:

$$AMAE_{R_j} = \frac{1}{P_{R_j}^2} \sum_{i=1}^{N} \frac{AE_i}{N} \qquad i \in R_j, \quad j = 1,2,\dots J \qquad (2)$$

where:

$$AE_i = \begin{cases} |\hat{y}_i - y_i| & if\ |\hat{y}_i - y_i| < 183 \\ 365 - |\hat{y}_i - y_i| & if\ |\hat{y}_i - y_i| \geq 183 \end{cases} \qquad (3)$$

and where $AMAE_{R_j}$ is the adjusted mean absolute error for region R_j, N is the number of stations located in hydro-climate class R_j, $\hat{y}_i$ is the prediction while y_i is the equivalent observation of flood timing for a specific site i within hydro-climate class R_j, and P_{R_j} is the proportion of locations in hydro-climate class R_j having available data for predictors (ranging from 0 to 1). This metric was used to penalise predictors that are unavailable for many locations within a specific hydro-climate class (the square value emphasises the importance of this metric). The predictor X_{R_j} and value of γ_{R_j} that minimised the error for a given climate class were selected for each terminal node of the prediction scheme.

The tree-based flood timing prediction model, calibrated to the seasonal flood timing observed across the selected stations, was then applied to all land grid cells of the ERA-Interim dataset to derive a map of flood timing. The temporal concentration (*R* value) of selected climate predictor at each grid cell was then used to represent the confidence of the prediction (high confidence: *R* value ranges from 0.8 to 1.0; medium confidence: *R* value ranges from 0.6 to below 0.8; low confidence: R below 0.6). This information is useful to indicate areas with complex temporal distribution of the most important predictors, which could reduce the usefulness of flood timing prediction. For example, locations with a bimodal distribution of the PD predictor may have intense rainfall events distributed in both April and November, but the averaged timing (used to predict flood timing) would fall in February. The confidence of flood timing prediction (i.e. the flood timing prediction of February) therefore would be low in these cases.

3 Results and discussion

3.1 Seasonality characteristics of flood at the global scale

Figure 5 provides an overview of flood seasonality at gauged locations at the global scale (regional maps provided in supplementary). Figure 5a illustrates the average timing of floods for the 1981-2010 period, while Figure 5b shows the flood timing concentration R. Stations that exhibit uniformity in the records are highlighted as red dots in the lower panel. There is a clear regional association in the timing of flood occurrence, of which the patterns over North America and Europe concur with prior studies [*Blöschl et al.*, 2017; *Burn and Whitfield*, 2016; *Hall and Blöschl*, 2018; *Villarini*, 2016]. The selected stations provide streamflow observation for 3,539 of the total 57,191 ERA-Interim cells (noting that there may be several streamflow gauging stations in a single ERA-Interim cell), leaving 94% of the global ERA-Interim landmass ungauged. Selected stations are also unevenly distributed, with the percentage of cells having flood data is relatively high over North America and Europe (17.5 and 12.5% of ERA-Interim cells respectively). South America and Oceania have 7.2 and 5.8% of the total land mass covered by streamflow gauge while Africa and Asia are covered by less than 1% of the continental total land mass.

Notwithstanding data coverage limitations, this analysis provides a first regional perspective of flood timing over parts of Asia (the majority of stations are located in Japan and India together with some stations available across Russia) and several regions in the southern hemisphere (the majority of stations are located in Brazil and Australia). In Asia, high latitude regions have floods occurred typically during spring while the rest of this continent is dominated by summer to autumn floods. In the southern hemisphere, there is a clear transition of flood timing in the latitudinal direction. Due to the limited availability of snowmelt processes in the southern hemisphere (only significant in some mountainous areas as discussed in Section 2.2.2), the rainfall regime and its interaction with catchment soil moisture conditions are more likely to be the key flood generation mechanisms across these regions.

The strength of the seasonal cycle (Figure 5b) demonstrates a high level of spatial heterogeneity. There are several clusters of stations showing uniformity due to the influence of climate-related processes that have been documented in previous studies. For instance, the east of U.S. is subject to a range of flood generation processes occurring throughout the year such as tropical and extratropical storms, or snowmelt dynamics [*Villarini and Smith*, 2010]. European stations located at the foothill of mountainous areas tend to be influenced by a mix of spring-snowmelt, rainfall events and/or glacier melting in summer [*Hall and Blöschl*, 2018]. The southern coast of south-eastern Australia has frequent rainfall in winter, but heavier summer precipitation is also possible due to convective activity. The combined influence of extreme rainfall and antecedence soil moisture is a likely reason for uniformity in flood timing records across this region [*Leonard et al.*, 2008], particularly where soil moisture conditions are counter-cyclical with heavy rainfall (e.g. the most intense rainfall may occur during summer due to convective processes, but on average the soils tend to be wettest during the winter). Lastly, the south of Brazil is characterised by a non-defined rainy season due to the combined influence of cold fronts, thunderstorms, and tropical cyclones which make rainfall-induced floods occurring throughout the year [*Rao and Hada*, 1990; *Teixeira and Satyamurty*, 2011]. Ultimately of the 7,894 selected records, the uniformity hypothesis was rejected for 7,040 locations, and this subset of stations that exhibited significant seasonality in flood timing represents the final subset used for the prediction of flood timing.

3.2 Distribution of predictors with the least discrepancy to flood timing

The distribution of the ‘best climate predictor’ for the globe is provided in Figure 6 (regional maps for areas with a high density of stations are provided in supplementary Figure S3). An interesting pattern observed through this analysis is the high level of spatial clustering in the distribution of predictors having the least discrepancy to flood timing, suggesting the existence of homogeneous regions in terms of climate predictors that could be used to predict flood timing.

In regions above 35°N where snowmelt also plays a significant role in flood generation, there are clear regional patterns regarding the most important predictor of flood timing. In particular, snowmelt-dominant predictors (i.e. TD, SD and SD7 which usually occur in spring) are generally most suitable in the north-central and the north-east of the U.S., most of Canada, Central and North-Eastern Europe, North Eurasia, and Scandinavia. On the other hand, the rainfall-dominant predictors (i.e. PD, PD7, PD30, and PD90) are generally the most suitable to explain flood occurrences on the western coastline of North America and Western Europe (including the UK). These findings are generally consistent with previous studies [*Berghuijs et al.*, 2016; *Burn and Whitfield*, 2016; *Cunderlik and Ouarda*, 2009; *Hall et al.*, 2014; *Mediero et al.*, 2015; *Villarini and Smith*, 2010; *Ye et al.*, 2017].

Focusing on regions with no snowmelt-based predictors (i.e. below 35°N), short-term precipitation predictors (PD and PD7) generally have the closest match with the timing of floods in the south-eastern US, northern Australia, and both the eastern and southern regions of Brazil, where previous studies have shown the importance of thunderstorm activities or tropical cyclones in flood generation [*Ávila et al.*, 2016; *Bradley and Smith*, 1994; *Stevenson and Schumacher*, 2014; *Villarini*, 2016; *Villarini et al.*, 2014]. On the other hand, long-term precipitation predictors (PD30 and PD90) have the highest consistency with flood timing in central Brazil and southern Australia, while other regions show a mixture between these two groups.

This comparison shows two of the main challenges for predicting flood timing at the global scale. Firstly, within relatively small geographic areas, such as the US Rocky Mountains or the Alpine region in Europe, there is large variability in the identified predictor, which reflects the complexity of flood formation factors (snowmelt, soil moisture state of the catchment, and different types of precipitation) across these regions [*Berghuijs et al.*, 2016; *Parajka et al.*, 2010]. Secondly, many locations also show a high correlation between predictors (e.g. the average timing of short-term precipitation and long-term precipitation being in the same month, see supplementary Figure S4), and this feature creates noise in determining the most important predictor. In addition, it also indicates a limitation of the prediction scheme, as the dependences between short-term precipitation and long-term precipitation predictors cannot be fully reflected (e.g. the single most extreme rainfall event may occur at the end of rainfall season and thus PD and PD90 have the similar values). Nevertheless, the spatial patterns shown in Figure 6 indicate the utility of the climate predictors to identify different flood-timing mechanisms at the regional scale.

The level of consistency between flood timing and available predictors (i.e. the discrepancy, in number of days, between flood timing and available predictors as defined in Table 1) was also analysed to evaluate the appropriateness of using these predictors for estimating flood timing. At the continental scale (Figure 7), all precipitation-based predictors generally have a good level of consistency in Asia, Africa, and South America, with more than 70% of stations exhibiting high or medium consistency with flood timing. In Oceania (of which the majority of stations are in Australia), flood timing is most consistent with long-term precipitation predictors, as both PD30 and PD90 have more than 60% of stations exhibiting high or medium consistency. In North America and Europe, where snowmelt-related processes are a key flood-producing mechanism, the percentage of stations showing high or medium consistency between precipitation-based predictors and flood timing is lower than the other continental regions; however, this is supplemented by snowmelt predictors, which have high and medium consistency for approximately 25-40% of stations.

The level of consistency between flood timing and the single most important predictor across the 7,040 stations was also assessed (showed in Table 3), suggesting generally high consistency at the global scale with the percentage of stations having high and medium levels of consistency being 50.9% and 31.8% respectively. This pattern is also evident at the continental level, with the percentage of locations showing high or medium consistency levels ranging from 72% (Oceania) to 97% (Africa). These results indicate the potential of using the proposed indices to predict flood timing, which could result in a model with up to 80% of locations having a prediction error of less than 46 days (i.e. the predicted and observed flood timing will fall within the same season).

3.3 A hydro-climate classification to estimate global flood timing

A rule-based classification (Figure 8b; herein referred to as D5) was developed to partition the land surface into five hydro-climate classes (Figure 8a). Although it is possible to further break each class into sub-regions and potentially improve the model's predictive power, the classification scheme was kept at this level of simplicity because the tree is found to represent the key regional patterns of the best predictors. In addition, the high correlation between predictors within the same group (e.g. PD30 and PD90; see Supplementary Figure S4) indicates that breaking these classes into sub-classes does not necessarily lead to improved accuracy in terms of predicting flood timing. Among the 11 'candidate' separating variables, four were retained for the final classification (MAP, P_{swet}, $TD_{indicator}$, and $f_{snow,}$), which partition the world into three rainfall-dominant classes (Class 1 to Class 3) and two snowmelt-dominant classes (Class 4 and Class 5).

As shown in the resulting tree, the first splitting rule focuses on differentiating rainfall-dominant classes from snowmelt-dominant classes. Specifically, the index $TD_{indicator}$ was used as the splitting variable, reflecting the fact that regions where the transition timing predictor (*TD*) cannot be reliably defined (i.e. $TD_{indicator}$ = 0) are unlikely to have snowmelt occurring. These "no snow-melt" regions were then divided into two classes using the total amount of annual precipitation. Specifically, locations with annual precipitation higher than 1200mm (or higher than annual rainfall of approximately 80% of all land grid cells) were assigned into Class 1 while the other locations were assigned into Class 2. Locations satisfying this condition (i.e. *MAP* >

1200mm) are mostly coastal areas or tropical regions (see Figure C1 in Supplementary), and are often characterised by strong activity of thunderstorms and tropical cyclones. Class 1 is therefore more likely to have short-term precipitation driving floods relative to Class 2.

For locations where the *TD* predictor can be reliably estimated, the "transitional regions" between rainfall-dominant and snowmelt-dominant groups were identified using the fraction of precipitation falling as snow (f_{snow} < 0.2). The key characteristic of these "transitional regions" is a relatively low amount of snowfall (and thus snowmelt) occurring, so rainfall mechanisms may still play a dominant role in flood generation. Across these "transitional regions", rainfall-dominant locations (Class 3) were defined if more than 12% of precipitation falls into the wettest month of spring-summer period (i.e. P_{swet}/MAP > 0.12), while the other locations were classified as snowmelt-dominant (i.e. Class 4). This splitting rule suggests that locations where rainfall concentrates in a specific month may potentially have floods that are driven by rainfall processes. The final class of the prediction scheme (Class 5) is characterized by a higher fraction of precipitation fall as snow ($f_{snow} \geq$ 0.2), and thus floods are more likely driven by snowmelt processes.

The dominant atmospheric predictor of flood timing was then identified for each hydro-climate class to form a linear function between that predictor and the flood timing response. The most relevant predictor and associated lag-day in each class were identified through the optimisation process described in Section 2.3.2 and are presented in Table 4. Although this process was automated, the chosen predictors are generally consistent with the splitting rules determining the boundaries. Among 7,040 locations, the prediction scheme could be applied for 6,671 stations in total (excluding 369 stations due to a missing value of the identified predictor). The majority of 'no prediction' locations fall into Class 1 due to the non-seasonal characteristic of rainfall-predictors across the south-eastern U.S., which contains most stations classified into Class 1. The maximum value of γ_{R_j} across five hydro-climate classes was found to be 15 days, indicating that floods, on average, occur within the 15-day window from the timing of the dominant predictor. Prediction errors (represented by mean absolute error) range from 21 days (Class 1) to 34 days (Class 3), and when averaged across all stations had a value of 31 days. Across all land regions, snowmelt, long-term precipitation and short-term precipitation predictors respectively predict flood timing for 43.3%, 29.1% and 27.6% of the global landmass.

Although the overall performance of the prediction scheme is reasonable at the global scale, there are some regions that have a large prediction error (Figure 9) such as central North America or the Alps (regional maps provided in Figure S5 of the supplementary material). There were many locations within these regions that exhibited non-seasonality in flood timing (e.g. central North America or the Alps; reported in Section 3.1), indicating some limitations in the proposed prediction scheme, which will be further discussed in our "caveats" section.

The global prediction of flood timing using the proposed classification system (Figure 10a), however, can reflect most of the large-scale spatial association in flood timing, especially in the southern hemisphere, where rainfall plays the key role in flood generation. The longitudinal transition over regions with high station density (e.g. North America and Europe) is also generally illustrated, suggesting the potential capacity of this prediction scheme in representing the spatial complexity of flood generation processes. The prediction of flood timing not only has consistency with flood timing based on regional observational studies in Europe and North America, but also has high consistency with the spatial patterns of the main high-flow season obtained from a global hydrological model [*Lee et al.*, 2015]. Additionally, the predicted flood timing is compared favourably to the streamflow peak month identified monthly stream flow series across 1,345 sites globally [*Dettinger and Diaz*, 2000] and the recently published GRUN gridded runoff product [*Ghiggi et al.*, 2019], providing confidence that a relatively simple predictive scheme—based on readily available atmospheric predictors obtained from reanalysis datasets—is able to provide credible predictions of flood timing in both data rich and sparse regions.

Figure 10b illustrates the prediction confidence base on the temporal concentration (*R* value) of selected predictor across the globe. High latitude area and regions where floods are influenced primarily by intense rainfall events (e.g. south and south-east Asia) generally possess high to medium prediction confidence (i.e. selected predictor has *R* value higher than or equal to 0.6). The distribution of areas exhibiting low prediction confidence is quite consistent with the empirical assessment presented in Section 3.1. Specifically, the majority of low confidence prediction falls over arid areas (e.g. northern Africa, inland of Australia) or locations where there are no strong signal of the seasonal cycle of defined predictors (e.g. southern of Australia, south-eastern US).

3.4 Caveats of the proposed flood timing prediction scheme

The proposed prediction scheme, although possessed the capacity to reflect many important spatial association in flood timing, has two important caveats that should be taken into account for any considered application. The consistency analysis between flood timing and the predictors (e.g. short-term rainfall) assumes that flood, on average, would occur within a small time-window of the averaged timing of the most relevant hypothesised process. For simplicity, a common approach was applied across all the predictors, and thus does not consider whether the predictor occurs before or after the flood event. This is likely to be appropriate for long-term rainfall predictors (e.g. PD90), such that it is physically plausible for a flood to be caused by accumulated wetness yet having the averaged timing occurs before the predictor. This assumption is less physically realistic for shorter-duration (i.e. 'heavy rainfall' predictors - PD or PD7) in which one would expect the heavy rainfall to occur prior to the flood event. In addition, there is a possibility of significant co-linearity between predictors, implying that the annual maximum streamflow may have a close association between both short-term and long-term rainfall predictors. As a result, findings of predictor-flood timing relationships cannot be interpreted as a definitive statement of causality regarding the flood generation mechanisms for individual sites.

Another caveat of the proposed prediction scheme lies in the data-driven approach of the hydro-climate classification scheme. Specifically, the global prediction is primarily based on the analysis of the predictors with the least discrepancy to flood timing. This approach is generally sensitive to the climate datasets being used. For instance, using other reanalysis products such as ERA5 [*C3S*, 2017] or GSWP3 [*Kim*, 2017] could lead to some difference in the global map of flood timing.

The regions with large prediction errors (Figure 9) also indicates other shortcomings of the proposed prediction scheme, in which the data-driven approach may not correctly define the most important flood generation mechanism. This limitation is likely to occur over some relatively small geographic areas with large variability of the identified predictor, potentially in part due to the coarse resolution of climate reanalysis products. In addition, using a single most important predictor may not reflect the complexity in regions with more than one mechanism contributing substantially to flood generation. For example, flood timing across the Alps and the central North America is characterised with a multi-modal distribution (e.g. snowmelt dominant flood in spring and convective storms in summer) but only either snowmelt predictors (for the Alps) or rainfall predictors (for central North America) were chosen to predict flood timing.

4 Summary and conclusions

This study analysed the spatial consistency of observed flood seasonality from 7,894 streamflow records [*Do et al.*, 2018b; *Gudmundsson et al.*, 2018b] and climate variables derived from an atmospheric forcing reanalysis dataset [*Dee et al.*, 2011]. The analysis has not only demonstrated consistent results with existing studies of flood seasonality across Europe and North America [*Blöschl et al.*, 2017; *Burn and Whitfield*, 2016; *Hall and Blöschl*, 2018; *Villarini*, 2016], but has facilitated the extension of flood timing estimates across the globe. Having identified spatial consistency between flood timing and selected variables representing flood generating mechanisms, this study provides important observation-based evidence of homogeneous regions of flood generation mechanisms. Short-term precipitation predictors are highly correlated with flood timing in the south-eastern region of the US, northern Australia, and the southern and eastern regions of Brazil; long-term precipitation predictors are more relevant in central Brazil, western Europe, and southern Australia; and snowmelt predictors are the most important variables in the high-latitude areas of the North American and Eurasian continents. These findings complement current understanding of the average timing and temporal concentration of the maximum events, which is generally available for only North America and Europe. Stream-gauge scarcity remains the key limitation for gauge-based hydrological investigations at the global scale, with approximately 94% of the global landmass was not observed.

Notwithstanding the complexity of dominant flood producing mechanisms and data limitation, this study was able to empirically identify a low discrepancy between flood timing and a single most important atmospheric predictor over data-covered regions. The empirical analysis yielded high percentage of locations with discrepancy of less than or equal to 45 days; i.e. flood timing and the most suitable predictor occur in the same season (continental scale: 73% – 94%, global average 82%). Taking advantage of the strong agreement between flood timing and climate predictors, a rule-based classification system was developed to partition the world into five hydro-climate classes. Each class represents regions sharing a common flood timing predictor. The classification was used to infer flood timing globally, including regions not covered by streamflow gauges. Although there are some regions with a high prediction error (e.g. central North America, the Alps and southern Australia), the proposed model, which has a relatively simple structure, performs well in predicting flood timing (global mean absolute error of 31 days) and was able to preserve large-scale spatial associations in flood timing across the globe. The spatial pattern of flood seasons obtained from this analysis compares favourably to the high-flow seasonal data obtained from a global hydrological model [*Lee et al.*, 2015] and streamflow peak month obtained from 1345 sites globally [*Dettinger and Diaz*, 2000] or the recently published gridded runoff [*Ghiggi et al.*, 2019].

The classification system proposed in this study can be used to define regions of similar flood generation processes at the global scale. Considering its relative simplicity and reproducible character, the proposed prediction framework could also be used for different climate datasets to assess the variation in either flood timing or flood-generating processes. Finally, the global map of flood timing prediction could be used as a measure of global hydrological model performance, by providing an indicator that these models correctly simulate the climatic mechanisms that lead to large streamflow events.

Data and Acknowledgments

Observational streamflow index are taken from the GSIM archive and are freely available from http://dx.doi.org/10.1594/PANGAEA.887470 [*Gudmundsson et al.*, 2018a] and http://dx.doi.org/10.1594/PANGAEA.887477 [*Do et al.*, 2018a]. The authors thanks all the national agencies and institutions that made the streamflow data publicly available to be included in the GSIM archive. Gridded precipitation and temperature data are taken from ERA-Interim global atmospheric reanalysis and are available at https://www.ecmwf.int [*Dee et al.*, 2011]. Hong Xuan Do receives financial support from the Australia Award Scholarship (AAS), the D R Stranks Travelling Fellowship and is currently funded by School for Environment and Sustainability, University of Michigan through grant number U064474. This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

References

Alfieri, L., P. Burek, L. Feyen, and G. Forzieri (2015), Global warming increases the frequency of river floods in Europe, *Hydrol. Earth Syst. Sci.*, *19*(5), 2247-2260.

Asadieh, B., N. Y. Krakauer, and B. M. Fekete (2016), Historical Trends in Mean and Extreme Runoff and Streamflow Based on Observations and Climate Models, *Water*, *8*(5), 189.

Ávila, A., F. Justino, A. Wilson, D. Bromwich, and M. Amorim (2016), Recent precipitation trends, flash floods and landslides in southern Brazil, *Environmental Research Letters*, *11*(11), 114029.

Beck, H. E., A. de Roo, and A. I. J. M. van Dijk (2015), Global maps of streamflow characteristics based on observations from several thousand catchments, *Journal of Hydrometeorology*.

Bennett, B., M. Leonard, Y. Deng, and S. Westra (2018), An empirical investigation into the effect of antecedent precipitation on flood volume, *Journal of Hydrology*.

Berghuijs, W. R., R. A. Woods, C. J. Hutton, and M. Sivapalan (2016), Dominant flood generating mechanisms across the United States, *Geophysical Research Letters*, *43*(9), 4382-4390.

Berghuijs, W. R., S. Harrigan, P. Molnar, L. J. Slater, and J. W. Kirchner (2019), The relative importance of different flood-generating mechanisms across Europe, *0*(ja).

Beurton, S., and A. H. Thieken (2009), Seasonality of floods in Germany, *Hydrological Sciences Journal*, *54*(1), 62-76.

Bierkens, M. F. P. (2015), Global hydrology 2015: State, trends, and directions, *Water Resources Research*, *51*(7), 4923-4947.

Bierkens, M. F. P., et al. (2015), Hyper-resolution global hydrological modelling: what is next?, *Hydrological Processes*, *29*(2), 310-320.

Blöschl, G., et al. (2017), Changing climate shifts timing of European floods, *Science*, *357*(6351), 588.

Bradley, A. A., and J. A. Smith (1994), The hydrometeorological environment of extreme rainstorms in the southern plains of the United States, *Journal of Applied Meteorology*, *33*(12), 1418-1431.

Burn, D. H., and P. H. Whitfield (2016), Changes in floods and flood regimes in Canada, *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, *41*(1-2), 139-150.

C3S (2017), ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.

Cannon, A. J. (2012), Köppen versus the computer: comparing Köppen-Geiger and multivariate regression tree climate classifications in terms of climate homogeneity, *Hydrol. Earth Syst. Sci.*, *16*(1), 217-229.

Cunderlik, J. M., and T. B. M. J. Ouarda (2009), Trends in the timing and magnitude of floods in Canada, *Journal of Hydrology*, *375*(3–4), 471-480.

Cunderlik, J. M., T. B. M. J. Ouarda, and B. Bobée (2004), Determination of flood seasonality from hydrological records/Détermination de la saisonnalité des crues à partir de séries hydrologiques, *Hydrological Sciences Journal*, *49*(3).

Dankers, R., N. W. Arnell, D. B. Clark, P. D. Falloon, B. M. Fekete, S. N. Gosling, J. Heinke, H. Kim, Y. Masaki, and Y. Satoh (2014), First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble, *Proceedings of the National Academy of Sciences*, *111*(9), 3257-3261.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553-597.

Dettinger, M. D., and H. F. Diaz (2000), Global Characteristics of Stream Flow Seasonality and Variability, *Journal of Hydrometeorology*, *1*(4), 289-310.

Dhakal, N., S. Jain, A. Gray, M. Dandy, and E. Stancioff (2015), Nonstationarity in seasonality of extreme precipitation: A nonparametric circular statistical approach and its application, *Water Resources Research*, *51*(6), 4499-4515.

Do, H. X., S. Westra, and L. Michael (2017), A global-scale investigation of trends in annual maximum streamflow, *Journal of Hydrology*.

Do, H. X., L. Gudmundsson, M. Leonard, and S. Westra (2018a), The Global Streamflow Indices and Metadata Archive - Part 1: Station catalog and Catchment boundary, edited, PANGAEA.

Do, H. X., L. Gudmundsson, M. Leonard, and S. Westra (2018b), The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, *Earth Syst. Sci. Data*, *10*(2), 765-785.

Do, H. X., et al. (2019), Historical and future changes in global flood magnitude – evidence from a model-observation investigation, *Hydrol. Earth Syst. Sci. Discuss.*, *2019*, 1-31.

Ghiggi, G., V. Humphrey, S. I. Seneviratne, and L. Gudmundsson (2019), GRUN: An observations-based global gridded runoff dataset from 1902 to 2014, *Earth Syst. Sci. Data Discuss.*, *2019*, 1-32.

Gudmundsson, L., S. I. Seneviratne, and X. Zhang (2017), Anthropogenic climate change detected in European renewable freshwater resources, *Nature Climate Change*, *7*, 813.

Gudmundsson, L., H. X. Do, M. Leonard, and S. Westra (2018a), The Global Streamflow Indices and Metadata Archive (GSIM) - Part 2: Time Series Indices and Homogeneity Assessment, edited, PANGAEA.

Gudmundsson, L., H. X. Do, M. Leonard, and S. Westra (2018b), The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, *Earth Syst. Sci. Data*, *10*(2), 787-804.

Gudmundsson, L., M. Leonard, H. X. Do, S. Westra, and S. I. Seneviratne (2019), Observed trends in global indicators of mean and extreme streamflow, *Geophysical Research Letters*, *46*(2), 756-766.

Gudmundsson, L., et al. (2012), Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, *Journal of Hydrometeorology*, *13*(2), 604-620.

Hall, J., and G. Blöschl (2018), Spatial patterns and characteristics of flood seasonality in Europe, *Hydrol. Earth Syst. Sci.*, *22*(7), 3883-3901.

Hall, J., et al. (2014), Understanding flood regime changes in Europe: a state-of-the-art assessment, *Hydrol. Earth Syst. Sci.*, *18*(7), 2735-2772.

Hock, R. (2003), Temperature index melt modelling in mountain areas, *Journal of Hydrology*, *282*(1), 104-115.

Hodgkins, G. A., et al. (2017), Climate-driven variability in the occurrence of major floods across North America and Europe, *Journal of Hydrology*, *552*, 704-717.

Iliopoulou, T., et al. (2019), A large sample analysis of European rivers on seasonal river flow correlation and its physical drivers, *Hydrol. Earth Syst. Sci.*, *23*(1), 73-91.

Ingle Smith, D. (1999), Floods: physical processes and human impacts by K. Smith and R. Ward, John Wiley, Chichester 1998. No. of pages: 382, *Earth Surface Processes and Landforms*, *24*(13), 1261-1261.

Ishak, E., A. Rahman, S. Westra, A. Sharma, and G. Kuczera (2013), Evaluating the non-stationarity of Australian annual maximum flood, *Journal of Hydrology*, *494*, 134-145.

Ivancic, T., and S. Shaw (2015), Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge, *Climatic Change*, 1-13.

Johnson, F., C. J. White, A. van Dijk, M. Ekstrom, J. P. Evans, D. Jakob, A. S. Kiem, M. Leonard, A. Rouillard, and S. Westra (2016), Natural hazards in Australia: floods, *Climatic Change*, 1-15.

Kim, H. (2017), Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1), *Data Integration and Analysis System*.

Köppen, W. (1900), Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt, *Geographische Zeitschrift*, *6*(11), 593-611.

Kozlowski, T. T. (1984), CHAPTER 1 - Extent, Causes, and Impacts of Flooding, in *Flooding and Plant Growth*, edited by T. T. Kozlowski, pp. 1-7, Academic Press, San Diego.

Lee, D., P. Ward, and P. Block (2015), Defining high-flow seasons using temporal streamflow patterns from a global model, *Hydrol. Earth Syst. Sci.*, *19*(11), 4689-4705.

Leonard, M., A. Metcalfe, and M. Lambert (2008), Frequency analysis of rainfall and streamflow extremes accounting for seasonal and climatic partitions, *Journal of hydrology*, *348*(1-2), 135-147.

Mallakpour, I., and G. Villarini (2015), The changing nature of flooding across the central United States, *Nature Clim. Change*, *5*(3), 250-254.

Mardia, K. V., and P. E. Jupp (2009), *Directional statistics*, John Wiley & Sons.

Marengo, J. A. (2006), On the hydrological cycle of the Amazon Basin: A historical review and current state-of-the-art, *Revista Brasileira de Meteorologia*, *21*(3), 1-19.

Mediero, L., et al. (2015), Identification of coherent flood regions across Europe by using the longest streamflow records, *Journal of Hydrology*, *528*, 341-360.

Merz, B., N. V. Dung, H. Apel, L. Gerlitz, K. Schröter, E. Steirou, and S. Vorogushyn (2018), Spatial coherence of flood-rich and flood-poor periods across Germany, *Journal of Hydrology*, *559*, 813-826.

Parajka, J., et al. (2010), Seasonal characteristics of flood regimes across the Alpine–Carpathian range, *Journal of Hydrology*, *394*(1), 78-89.

Pathiraja, S., S. Westra, and A. Sharma (2012), Why continuous simulation? The role of antecedent moisture in design flood estimation, *Water Resources Research*, *48*(6), n/a-n/a.

Pilgrim, E., A. Institution of Engineers, D. Pilgrim, and R. Canterford (1987), *Australian rainfall and runoff*, Institution of Engineers, Australia.

Rao, V. B., and K. Hada (1990), Characteristics of rainfall over Brazil: Annual variations and connections with the Southern Oscillation, *Theoretical and Applied Climatology*, *42*(2), 81-91.

Sharma, A., C. Wasko, and D. P. Lettenmaier (2018), If Precipitation Extremes Are Increasing, Why Aren't Floods?, *Water Resources Research*, *0*(0).

Slater, L. J., and G. Villarini (2016), Recent trends in U.S. flood risk, *Geophysical Research Letters*, *43*(24), 12,428-412,436.

Stevens, A. J., D. Clarke, and R. J. Nicholls (2016), Trends in reported flooding in the UK: 1884–2013, *Hydrological Sciences Journal*, *61*(1), 50-63.

Stevenson, S. N., and R. S. Schumacher (2014), A 10-year survey of extreme rainfall events in the central and eastern United States using gridded multisensor precipitation analyses, *Monthly Weather Review*, *142*(9), 3147-3162.

Teixeira, M. d. S., and P. Satyamurty (2011), Trends in the Frequency of Intense Precipitation Events in Southern and Southeastern Brazil during 1960–2004, *Journal of Climate*, *24*(7), 1913-1921.

van Dijk, A. I. J. M., J. L. Peña-Arancibia, E. F. Wood, J. Sheffield, and H. E. Beck (2013), Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide, *Water Resources Research*, *49*(5), 2729-2746.

Villarini, G. (2016), On the seasonality of flooding across the continental United States, *Advances in Water Resources*, *87*, 80-91.

Villarini, G., and J. A. Smith (2010), Flood peak distributions for the eastern United States, *Water Resources Research*, *46*(6).

Villarini, G., R. Goska, J. A. Smith, and G. A. Vecchi (2014), North Atlantic tropical cyclones and US flooding, *Bulletin of the American Meteorological Society*, *95*(9), 1381-1388.

Ward, P. J., et al. (2015), Usefulness and limitations of global flood risk models, *Nature Clim. Change*, *5*(8), 712-715.

Wasko, C., and A. Sharma (2017), Global assessment of flood and storm extremes with increased temperatures, *Scientific Reports*, *7*(1), 7945.

Wasko, C., and R. Nathan (2019), Influence of changes in rainfall and soil moisture on trends in flooding, *Journal of Hydrology*, *575*, 432-441.

Westra, S., L. A. Alexander, and F. W. Zwiers (2013), Global Increasing Trends in Annual Maximum Daily Precipitation, *Journal of Climate*, *26*(11), 15.

Westra, S., H. J. Fowler, J. P. Evans, L. V. Alexander, P. Berg, F. Johnson, E. J. Kendon, G. Lenderink, and N. M. Roberts (2014), Future changes to the intensity and frequency of short-duration extreme rainfall, *Reviews of Geophysics*, *52*(3), 522-555.

Woldemeskel, F., and A. Sharma (2016), Should flood regimes change in a warming climate? The role of antecedent moisture conditions, *Geophysical Research Letters*, *43*(14), 7556-7563.

Wood, E. F., et al. (2011), Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resources Research*, *47*(5), n/a-n/a.

Woods, R. A. (2009), Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, *Advances in Water Resources*, *32*(10), 1465-1481.

Ye, S., H.-Y. Li, L. R. Leung, J. Guo, Q. Ran, Y. Demissie, and M. Sivapalan (2017), Understanding Flood Seasonality and Its Temporal Shifts within the Contiguous United States, *Journal of Hydrometeorology*, *18*(7), 1997-2009.

Tables

Table 1. Description of the five consistency categories between flood timing and a single predictor

Category	Description
High consistency	Discrepancy between (±)15 days
Medium consistency	Discrepancy between (±)16 and (±)45 days
Low consistency	Discrepancy between (±)46 and (±)75 days
Inconsistency	Discrepancy is outside of [-75, +75] range
No data available	Predictor data is not available at the reported coordinates of the streamflow station due to seasonal uniformity of the time series

Table 2. 'Candidate' climate indices for the rule-based hydro-climate classification.

Index	Description
MAP	mean annual precipitation (m)
MAT	mean annual temperature (^{o}C)
T_{hot}	temperature of the hottest month (^{o}C)
T_{cold}	temperature of the coldest month (^{o}C)
P_{dry}	precipitation of the driest month (m)
P_{sdry}	precipitation of the driest month in spring-summer$^{(*)}$ (m)
P_{wdry}	precipitation of the driest month in fall-winter$^{(*)}$ (m)
P_{swet}	precipitation of the wettest month in spring-summer$^{(*)}$ (m)
P_{wwet}	precipitation of the wettest month in fall-winter$^{(*)}$ (m)
f_{snow}	fraction of precipitation falling as snow (from 0 to 1). Daily precipitation is assumed to fall as rainfall when $T > 0$
$TD_{indicator}$	Binary variable (0/1) indicates whether transition time from snowfall to rainfall can be reliably identified (i.e. at least 70% of the years have a temperature rise from below to exceed 0^{o}C).

$^{(*)}$: spring-summer (fall-winter) is defined as the warmer (cooler) six-month period of October – March and April – September for each respective hemisphere.

Table 3. Number of stations grouped by five consistency categories at regional and global scales.

Continents	**Level of consistency**					**Total**
	High	*Medium*	*Low*	*Inconsistency*	*No data*	
Asia	223 (69.5%)	76 (23.7%)	15 (4.6%)	7 (2.2%)	0 (0%)	321
North America	1837 (44.7%)	1420 (34.5%)	509 (12.4%)	314 (7.6%)	31 (0.8%)	4111
Europe	703 (52.2%)	471 (35.0%)	95 (7.0%)	75 (5.6%)	3 (0.2%)	1347
Africa	100 (80.0%)	21 (16.8%)	2 (1.6%)	2 (1.6%)	0 (0%)	125
South America	544 (74.3%)	136 (18.6%)	26 (3.5%)	19 (2.6%)	7 (1.0%)	732
Oceania	177 (43.8%)	115 (28.5%)	44 (10.9%)	44 (10.9%)	24 (5.9%)	404
Global	**3584 (50.9%)**	**2239 (31.8%)**	**691 (9.8%)**	**461 (6.6%)**	**65 (0.9%)**	**7040 (100%)**

Table 4. Description of the hydro-climate classes defined through the D10 classification system (lower panel of Figure 6). 'No prediction' indicates locations where there is no predictor available to predict flood timing.

Class	Climate indices used to define hydro-climate class	Number of gauges	% of no prediction	Dominant flood generation	Lag-day	Prediction errors (MAE; in days)	% of global land mass
1	$TD_{indicator}$, MAP	1507	17	Short-term precipitation (PD7)	15	22	17.5
2	$TD_{indicator}$, MAP	1278	7	Long-term precipitation (PD90)	4	31	29.1
3	$TD_{indicator}$, f_{snow}, P_{swet}, MAP	709	3	Short-term precipitation (PD7)	-15	34	10.1
4	$TD_{indicator}$, f_{snow}, P_{swet}, MAP	2259	0	Snowmelt predictor (TD)	15	33	4.8
5	$TD_{indicator}$, f_{snow}	1287	1	Snowmelt predictor (SD7)	15	34	38.5
Global	$\boldsymbol{TD_{indicator}}$, $\boldsymbol{P_{swet}}$, $\boldsymbol{MAP}$, ***and*** $\boldsymbol{f_{snow}}$	**7,040**	**5**	**-**	**-**	**31**	**100.0**

Figures

Figure 1. Flow chart to make global prediction of flood timing using GSIM and ERA-Interim datasets.

Figure 2. Example of a station that does not have evidence to reject the null-hypothesis of uniformity in a circular time series (Fig. 2a; the east branch of Cann River located in Victoria, Australia), and a station that has evidence to reject the uniformity hypothesis (Fig. 2b; Los Sosa River located in Entre Rios Province, Argentina). Grey areas represent the density of maximum streamflow events distributed across 12 months of the year. The direction of the red arrow represents the average timing, whereas the length of the arrow illustrates the temporal concentration (R value) of the maximum events (0.1 and 0.9 for the Fig. 2a and Fig. 2b respectively).

Figure 3. Map of data availability for the seven predictors. Predictors were divided into two categories: (1) Rainfall-predictors comprising short-term rainfall predictors (PD and PD7) and long-term rainfall predictors (PD30 and PD90) and (2) Snowmelt-predictors comprising TD, SD and SD7. Unavailability may be due to no data being available (for snowmelt-base predictors only) or where the circular uniformity hypothesis was not rejected at the 10% level (for all predictors).

Figure 4. Illustration of the classification scheme and the procedure undertaken at each classification node.

Figure 5. Seasonality of flood occurrence across 7,894 GSIM stations fulfilling the quality control criteria for the period 1981-2010. Fig 5a: average flood timing; colour points represent long-term-mean value. Fig 5b: concentration index (R) of flood timing (values range from 0 to 1); red dots represent records with uniformity hypothesis was not rejected at the 10% significance (854 stations). In both panels: grey dots represent GSIM stations that were removed prior to this analysis due to quality entrance criteria (outlined in Section 2.1). Note that: (i) the averaged timing for points that are classified as 'uniform' would not be reliable; (ii) a low R may reflect a multi-modal distribution of flood timing, which is outside the scope of this study.

Figure 6. Global map of single predictor with smallest discrepancy to flood timing across 7,040 stations that exhibit seasonality in flood timing. The brown colours indicate the short-precipitation predictor (PD and PD7), blue colours represent the long-precipitation predictors (PD30 and PD90) and the red colours represent the snowmelt-base predictors (TD, SD and SD7). There are 63 stations with no data available for predictors. These stations are plotted in the grey colour.

Figure 7. Consistency between flood timing and individual predictors (top panels: snowmelt-based predictor; bottom panels: rainfall-based predictors), based on definitions in Table 1. Each bar chart illustrates the percentage of stations allocated into five consistency categories for one predictor across the six considered regions. Note that the top panels (TD, SD and SD7) have the same axis as the bottom panels.

Figure 8. Global maps of climate regions (top panel) partitioned by the D10 hydro-climate system (bottom panel). Each hydro-climate class is defined following a set of separation rules and has a prediction of flood timing as a linear function of one predictor.

Figure 9. Prediction errors across 7,040 stations grouped into the five consistency definitions in Table 1 based on local performance.

Figure 10. Global prediction of flood timing (Fig. 10a) and prediction confidence (Fig. 10b) using reanalysis climate forcing datasets and D10 decision tree. Grey colour indicates locations where there is no suitable predictor available due to lack of seasonality. Temporal concentration (*R*) of selected predictor was used to define prediction confidence for each cell.

Figure 1.

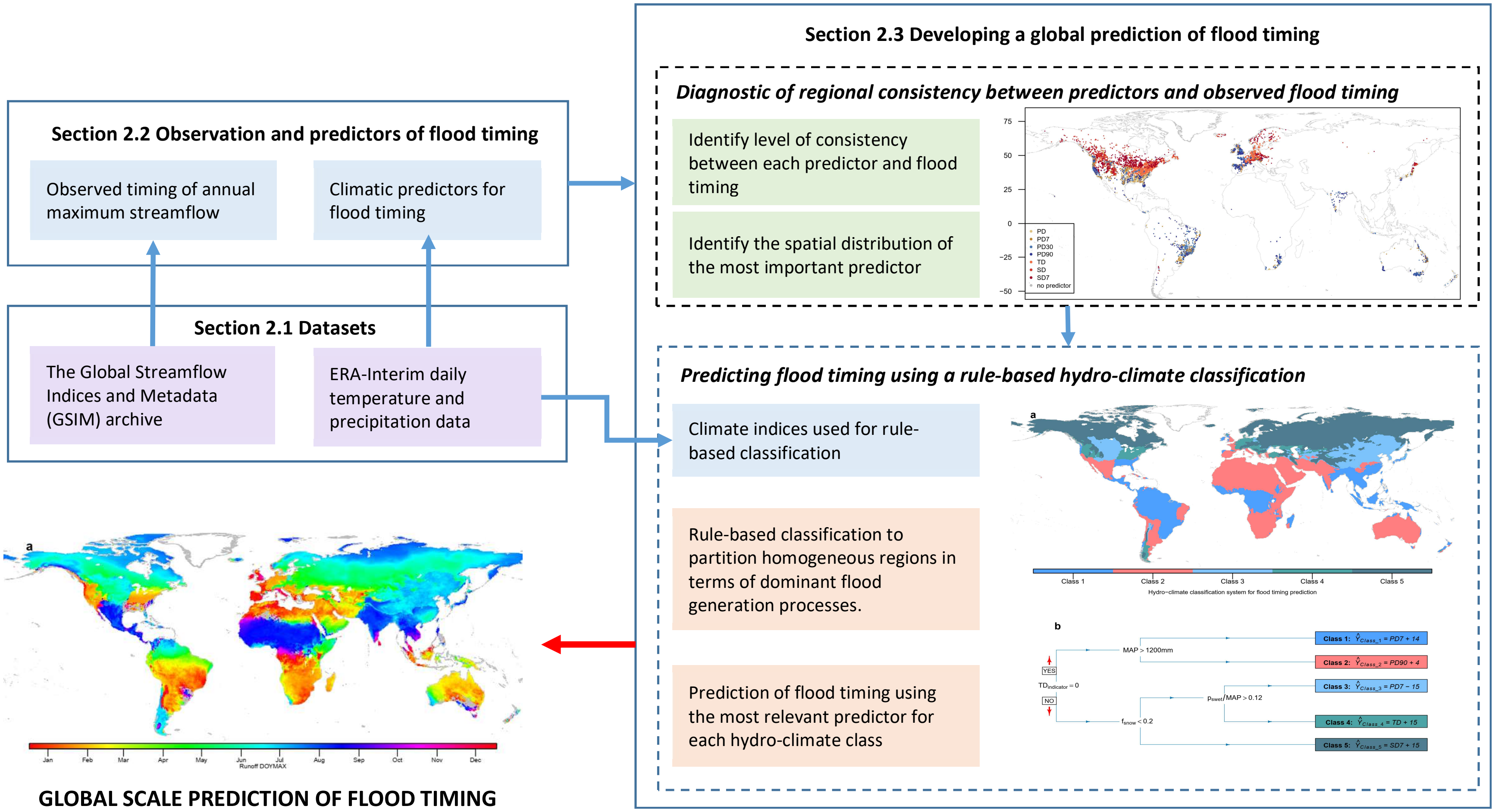

Section 2.2 Observation and predictors of flood timing
Observed timing of annual maximum streamflow
Climatic predictors for flood timing
Section 2.1 Datasets
The Global Streamflow Indices and Metadata (GSIM) archive
ERA-Interim daily temperature and precipitation data
Section 2.3 Developing a global prediction of flood timing
Diagnostic of regional consistency between predictors and observed flood timing
Identify level of consistency between each predictor and flood timing
Identify the spatial distribution of the most important predictor
PD
PD7
PD30
PD90
TD
SD
SD7
no predictor
Predicting flood timing using a rule-based hydro-climate classification
Climate indices used for rule-based classification
Rule-based classification to partition homogeneous regions in terms of dominant flood generation processes.
Prediction of flood timing using the most relevant predictor for each hydro-climate class
Class 1
Class 2
Class 3
Class 4
Class 5
Hydro−climate classification system for flood timing prediction
MAP > 1200mm
YES
NO
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Runoff DOYMAX
GLOBAL SCALE PREDICTION OF FLOOD TIMING

Figure 2.

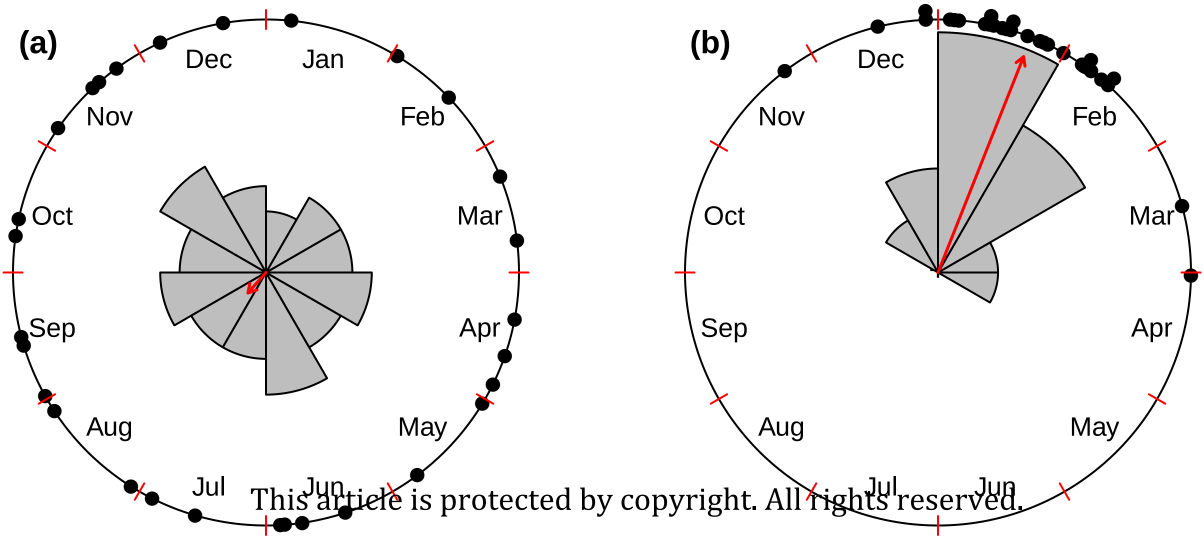

(a)
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
(b)
Dec
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov

Figure 3.

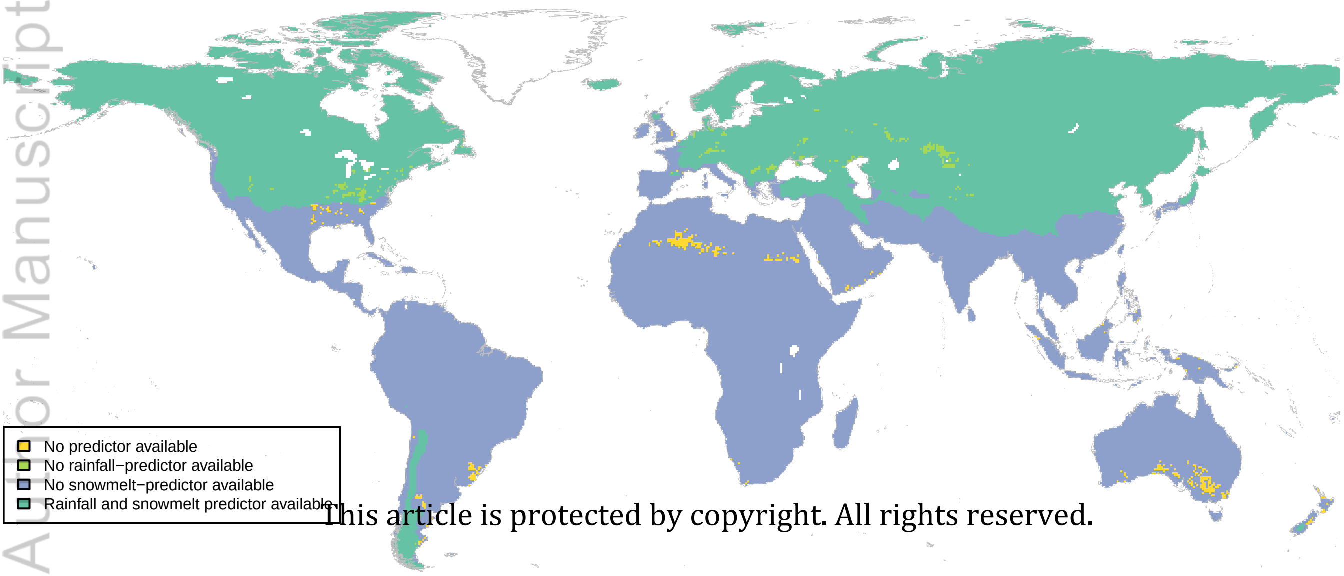

No predictor available
No rainfall−predictor available
No snowmelt−predictor available
Rainfall and snowmelt predictor available

Figure 4.

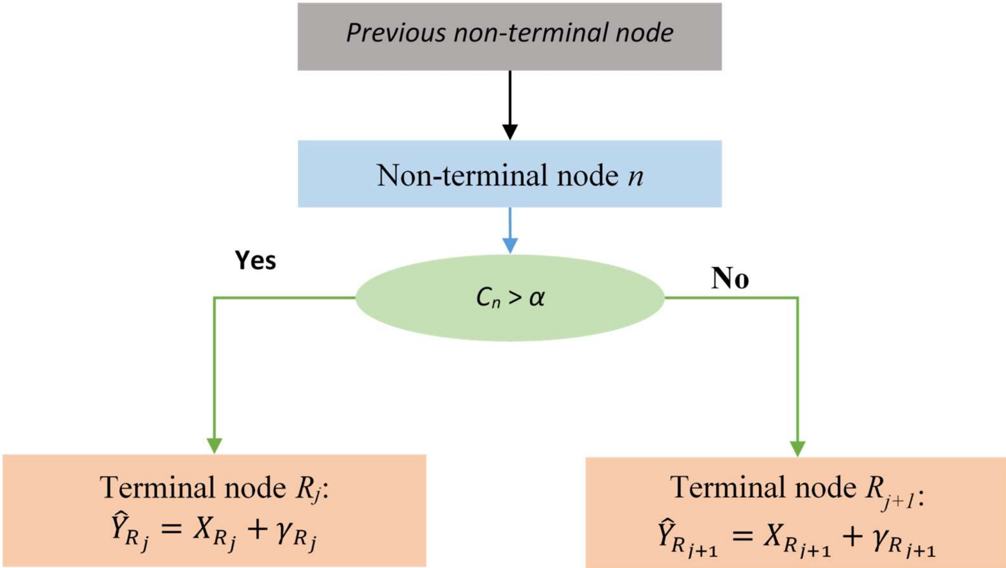
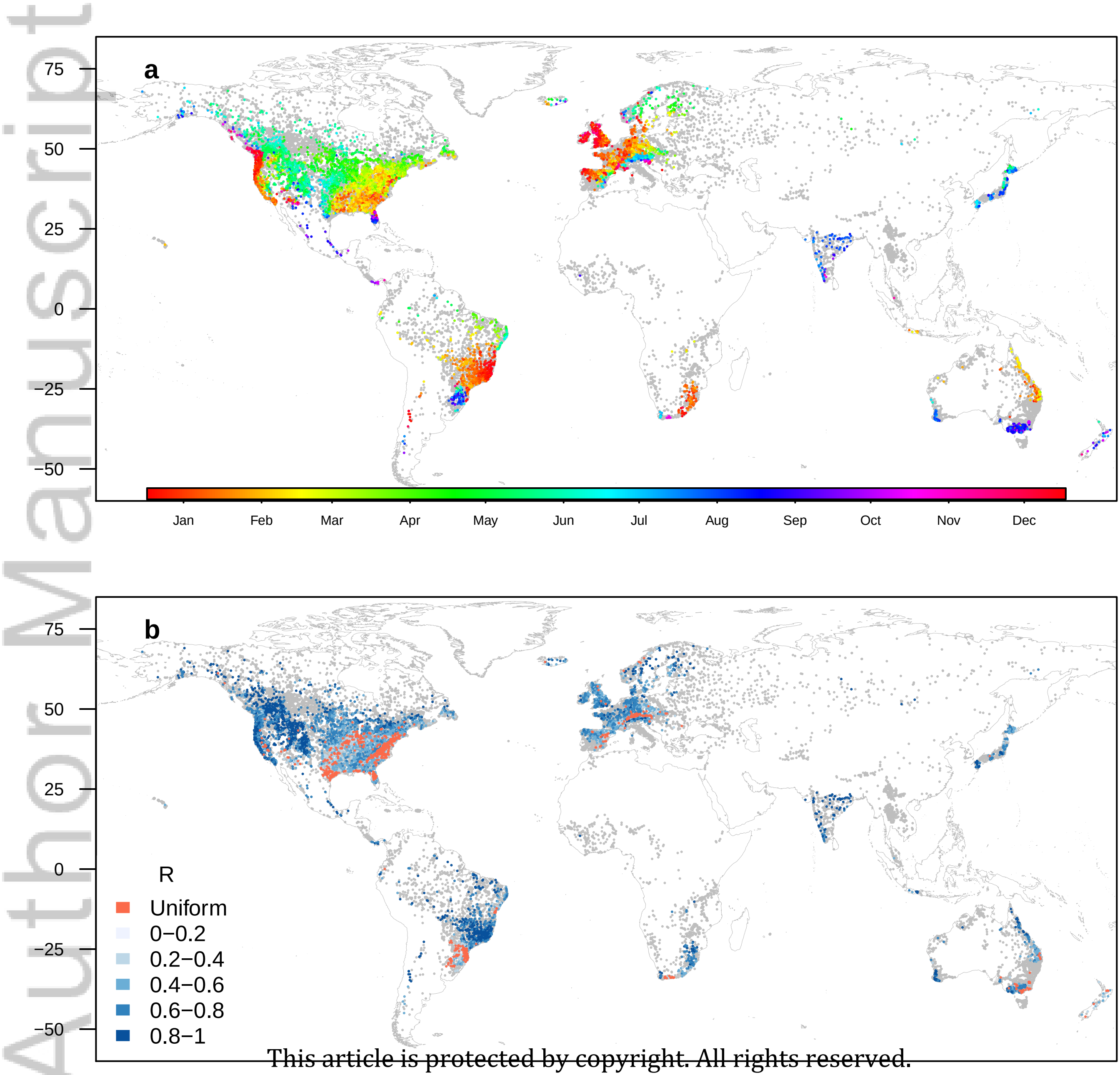

Previous non-terminal node
Non-terminal node n
Yes
$C_n > \alpha$
No
Terminal node R_j:
$\hat{Y}_{R_j} = X_{R_j} + \gamma_{R_j}$
Terminal node R_{j+1}:
$\hat{Y}_{R_{j+1}} = X_{R_{j+1}} + \gamma_{R_{j+1}}$

Figure 5.



a
75
50
25
0
−25
−50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
b
75
50
25
0
−25
−50
R
Uniform
0−0.2
0.2−0.4
0.4−0.6
0.6−0.8
0.8−1

Figure 6.

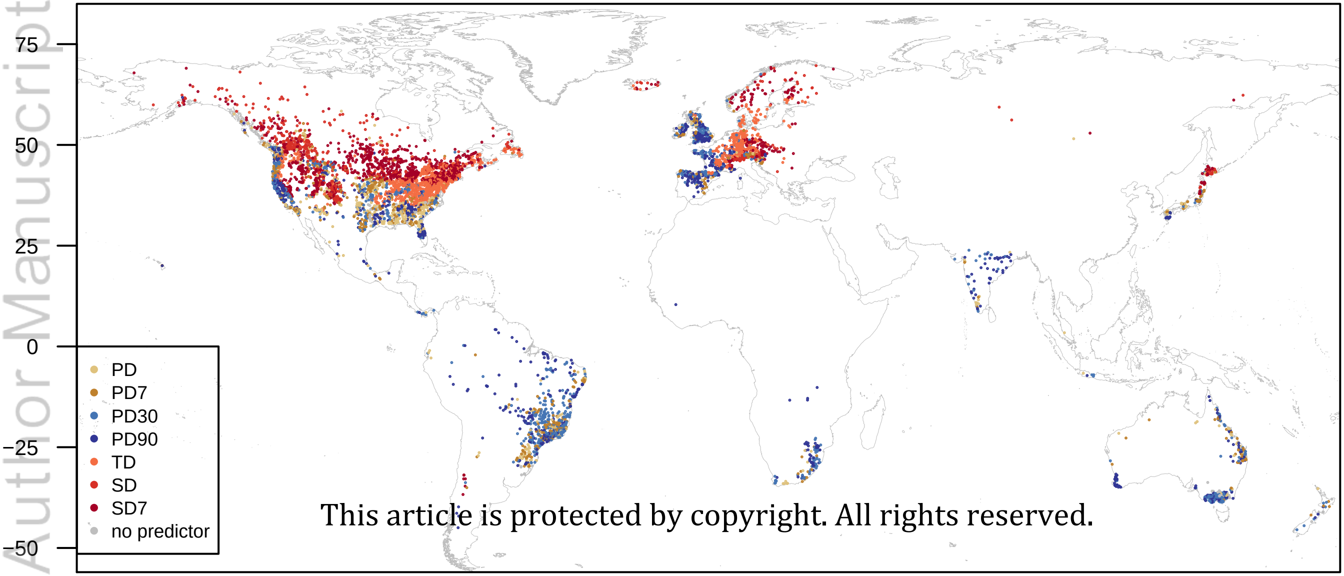

75
50
25
0
−25
−50
PD
PD7
PD30
PD90
TD
SD
SD7
no predictor

Figure 7.

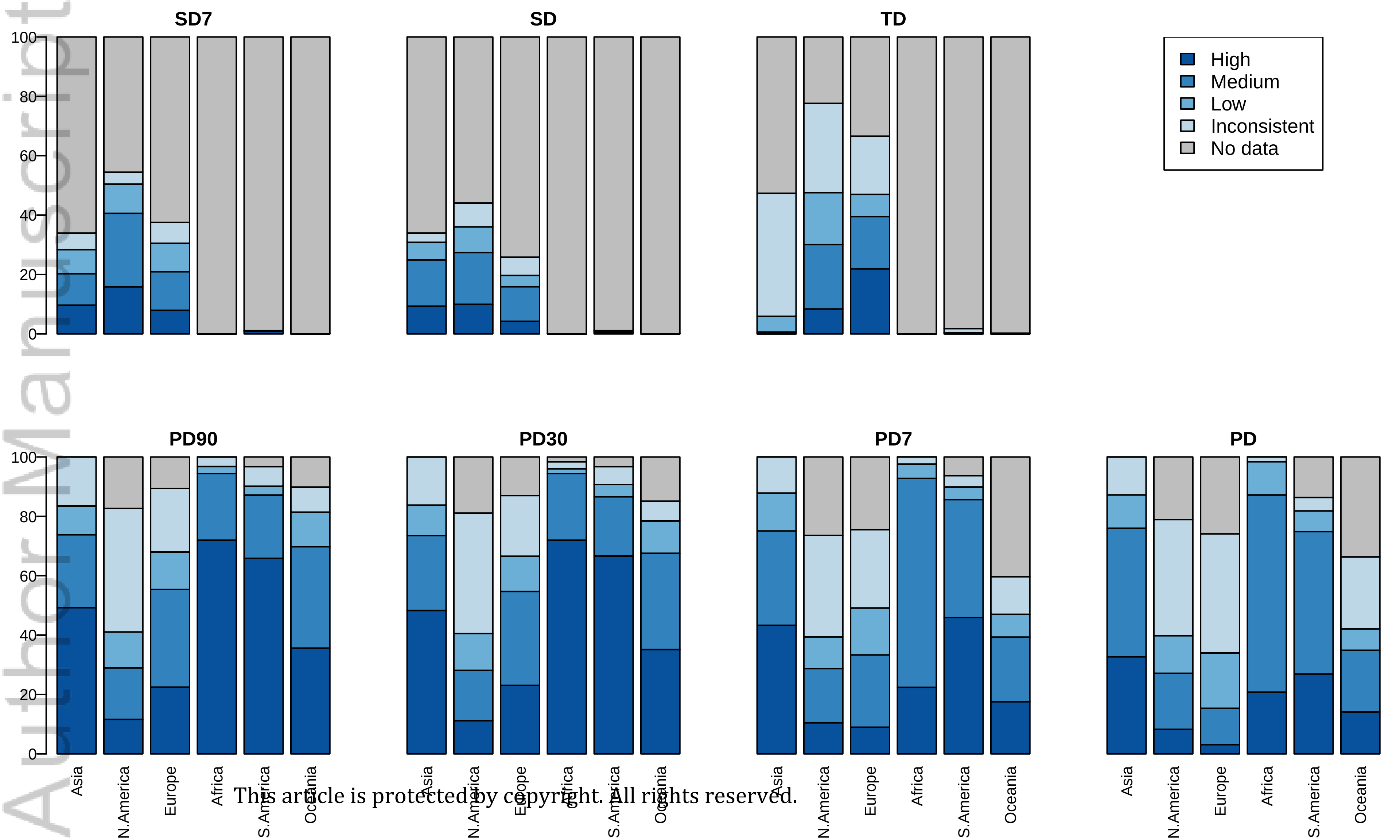

SD7
SD
TD
PD90
PD30
PD7
PD
High
Medium
Low
Inconsistent
No data
100
80
60
40
20
0
Asia
N.America
Europe
Africa
S.America
Oceania

Figure 8.

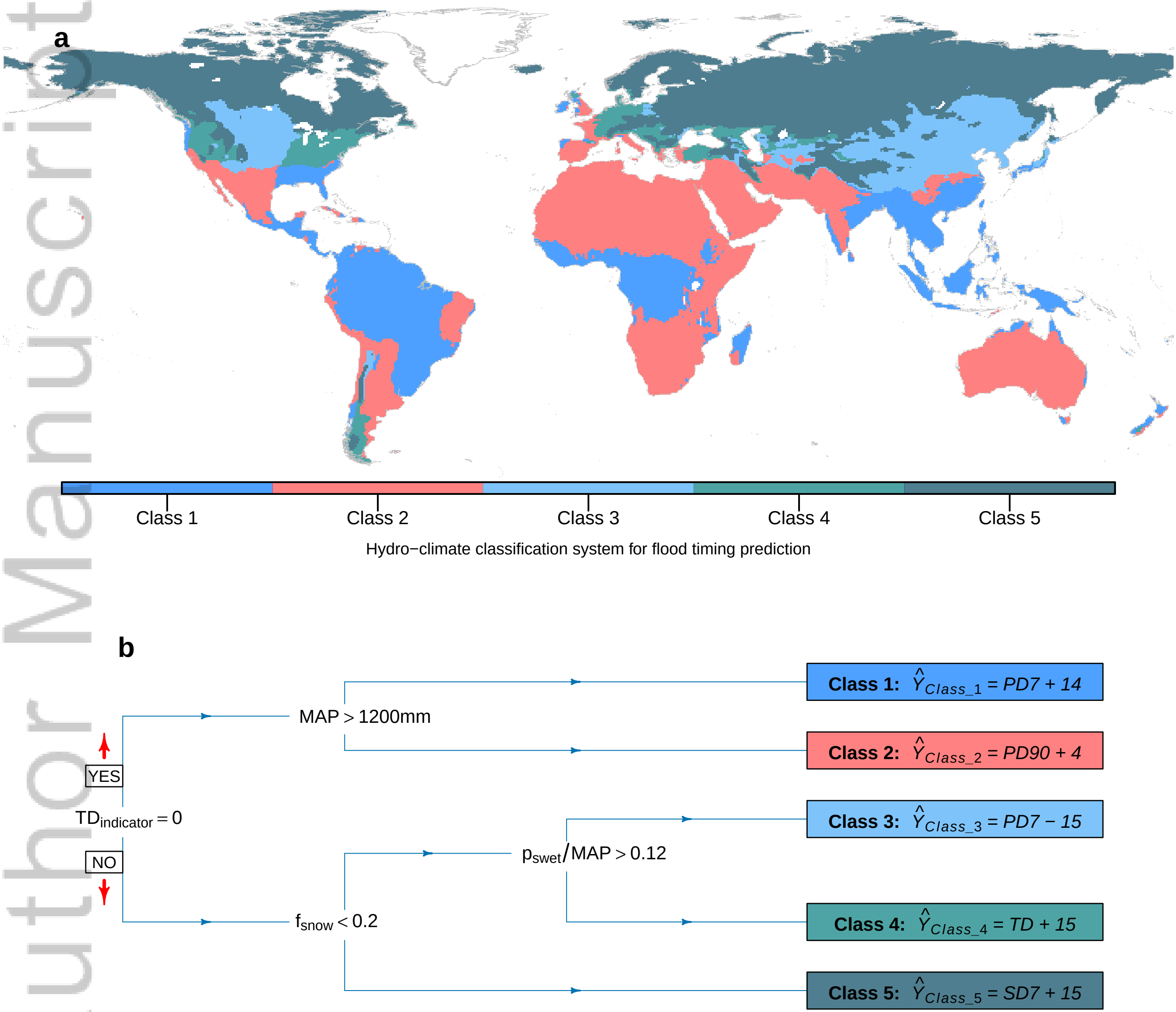

a
Class 1
Class 2
Class 3
Class 4
Class 5
Hydro−climate classification system for flood timing prediction
b
MAP > 1200mm
YES
TDindicator = 0
NO
fsnow < 0.2
pswet/MAP > 0.12
Class 1: Ŷ Class_1 = PD7 + 14
Class 2: Ŷ Class_2 = PD90 + 4
Class 3: Ŷ Class_3 = PD7 − 15
Class 4: Ŷ Class_4 = TD + 15
Class 5: Ŷ Class_5 = SD7 + 15

Figure 9.

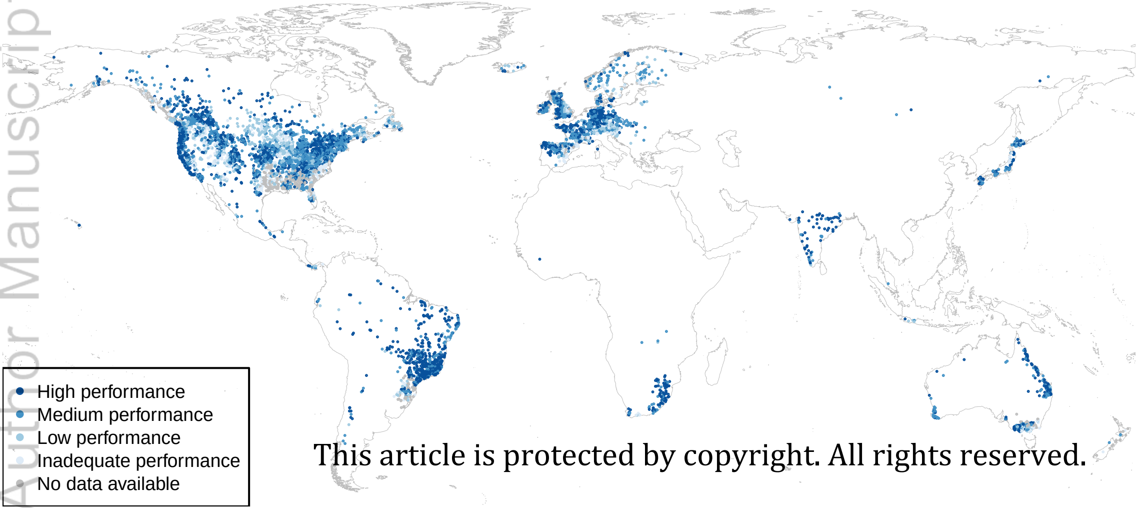
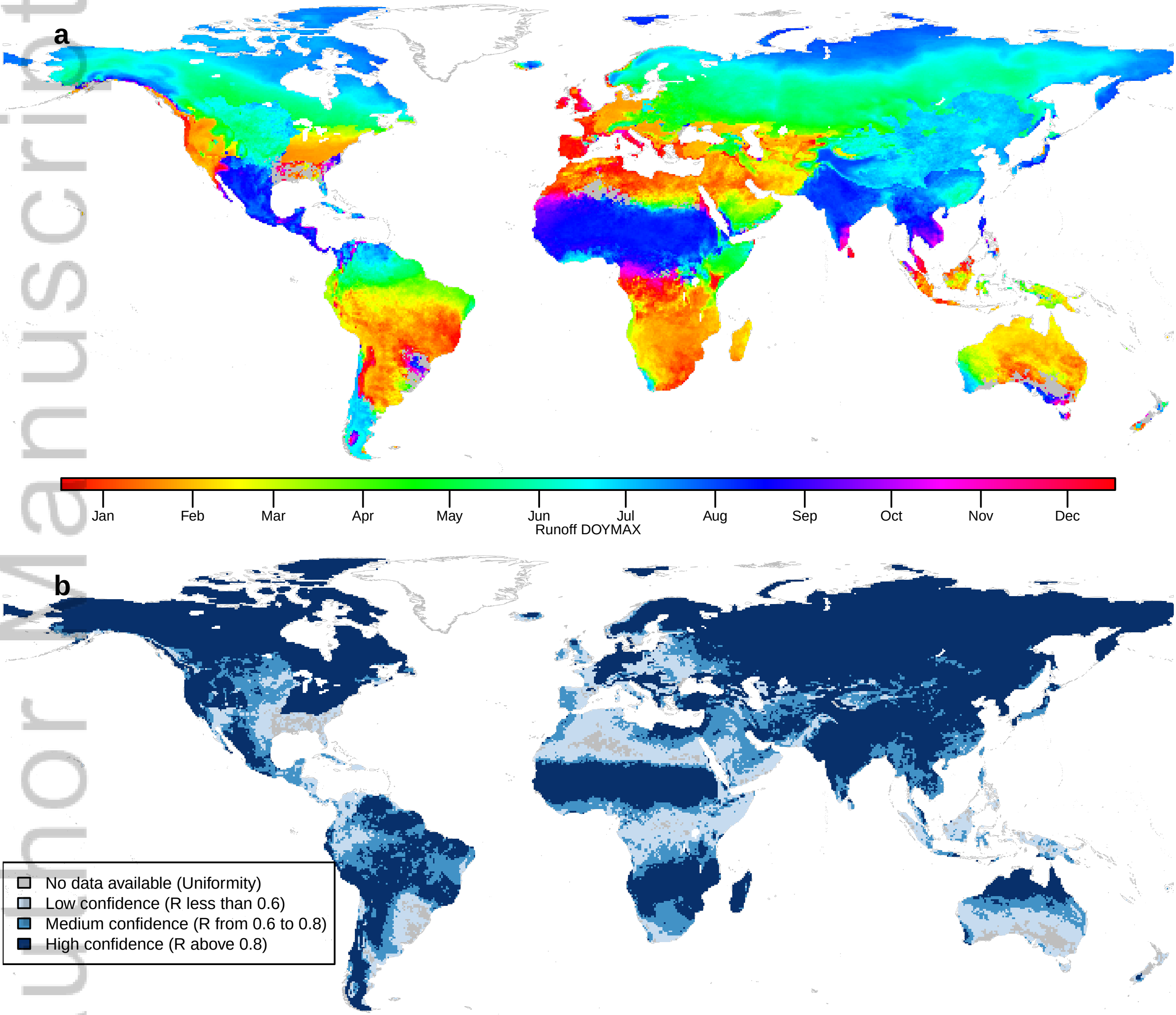
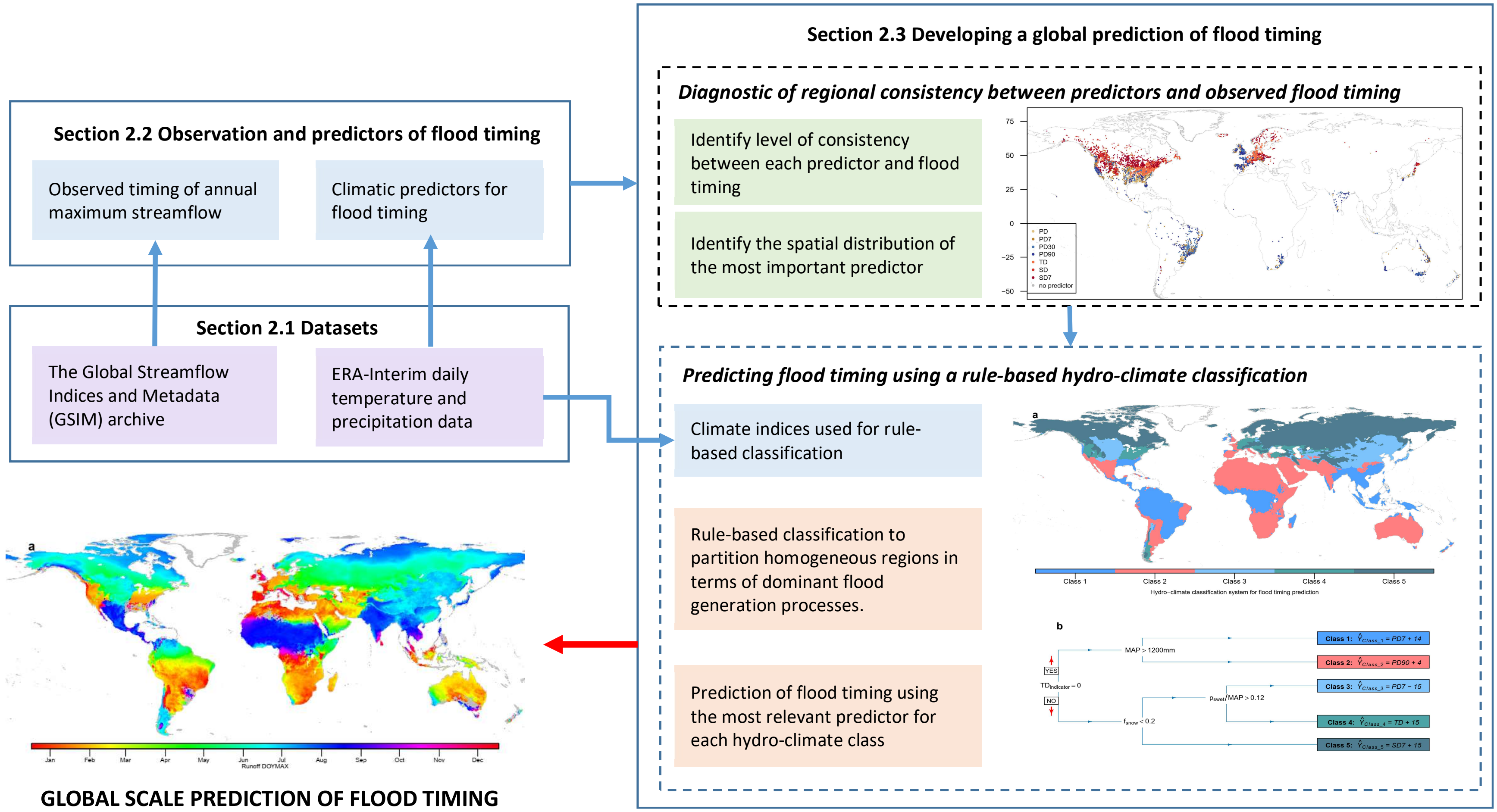
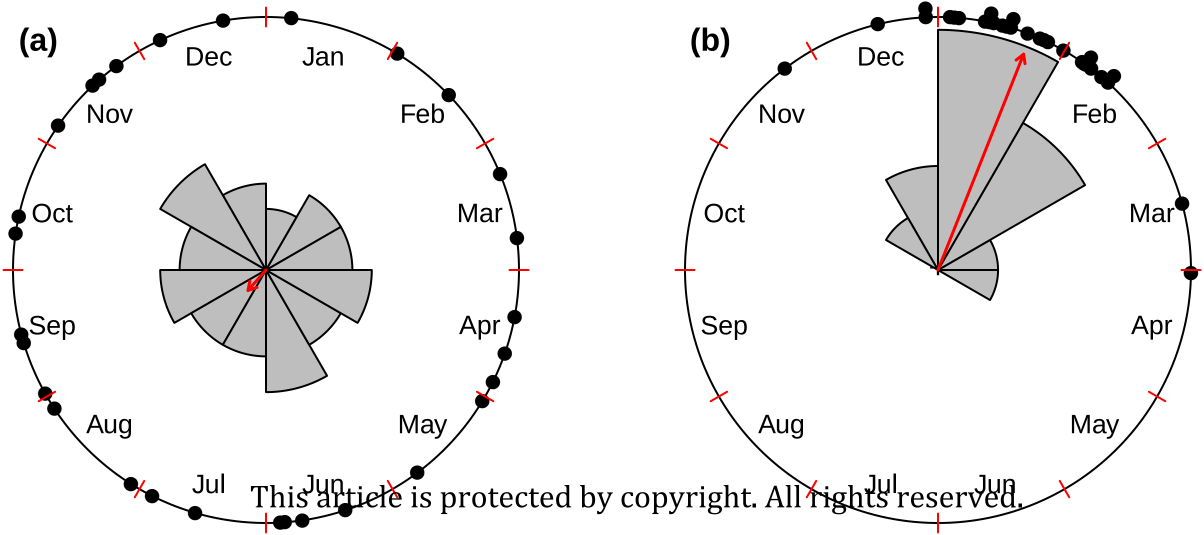
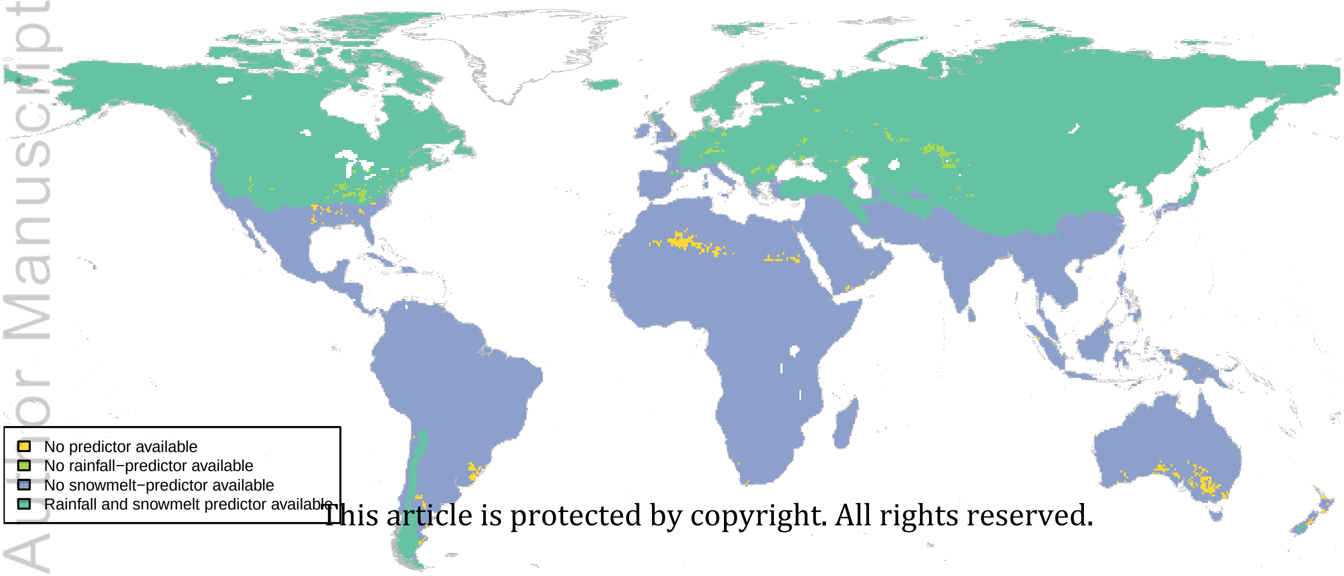

High performance
Medium performance
Low performance
Inadequate performance
No data available

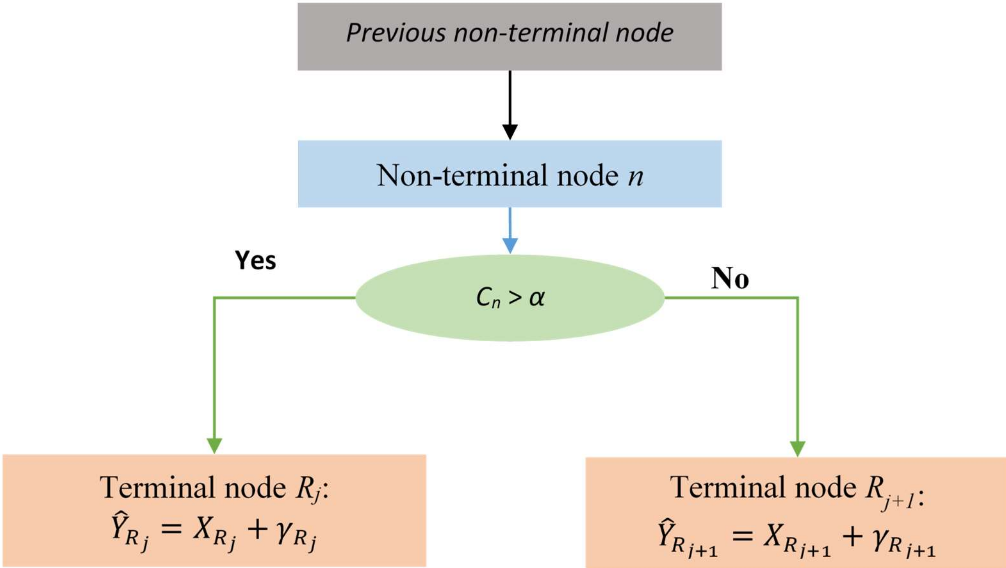
Figure 10.

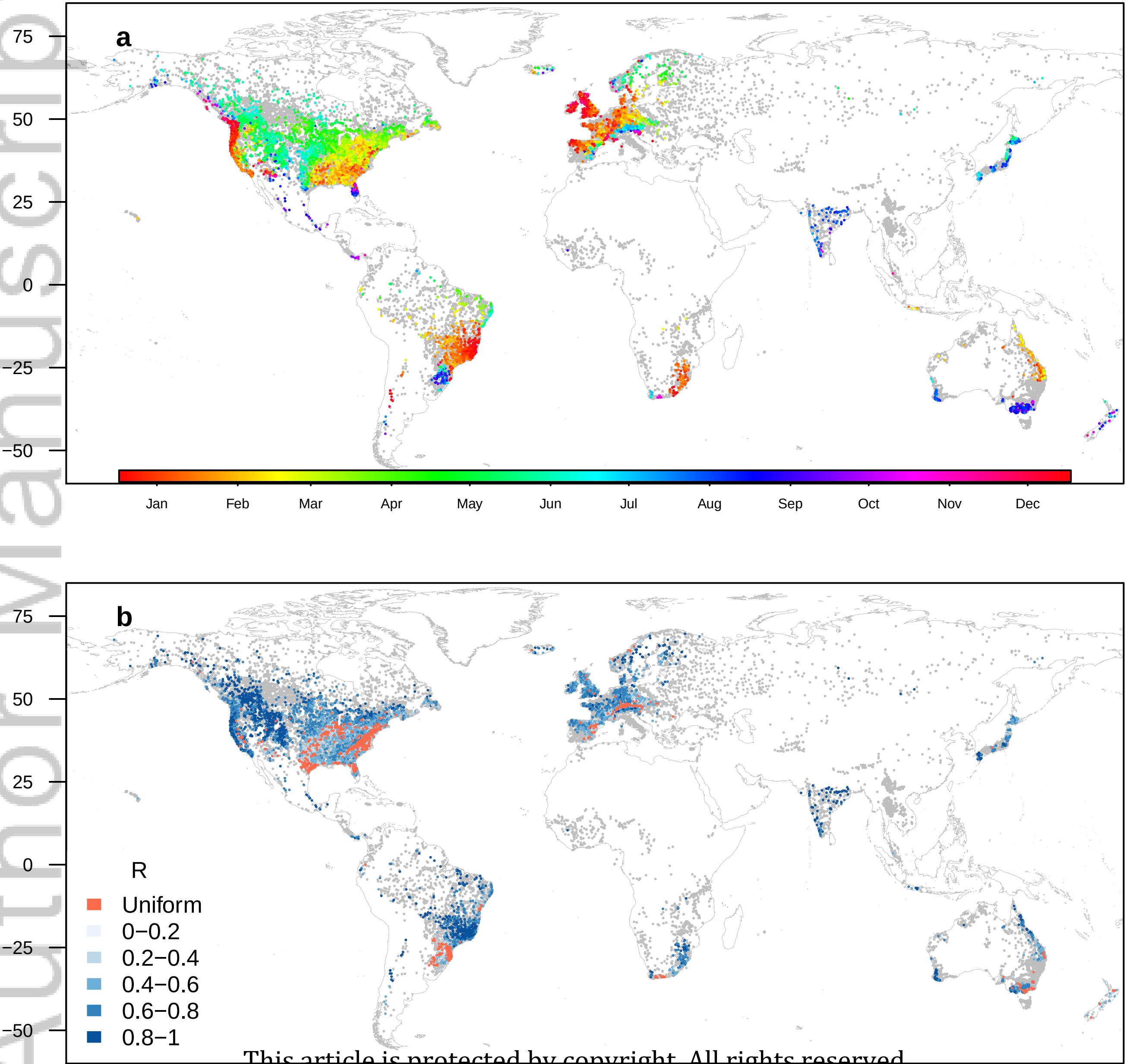

a
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Runoff DOYMAX
b
No data available (Uniformity)
Low confidence (R less than 0.6)
Medium confidence (R from 0.6 to 0.8)
High confidence (R above 0.8)

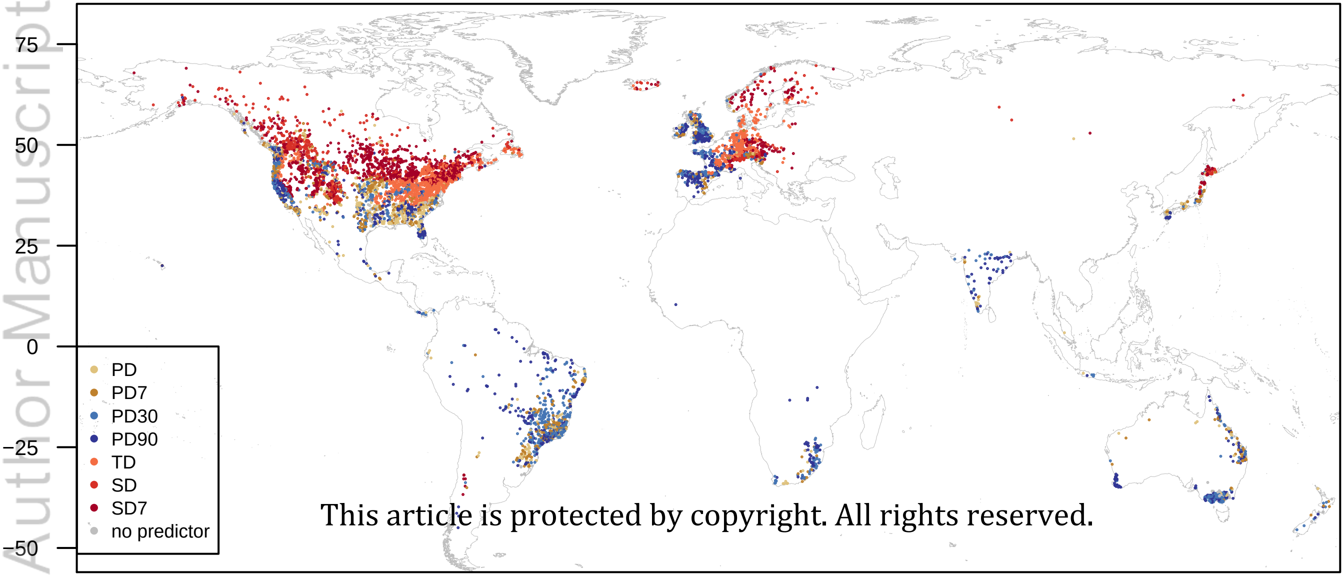

Section 2.3 Developing a global prediction of flood timing
Diagnostic of regional consistency between predictors and observed flood timing
Identify level of consistency between each predictor and flood timing
Identify the spatial distribution of the most important predictor
PD
PD7
PD30
PD90
TD
SD
SD7
no predictor
Predicting flood timing using a rule-based hydro-climate classification
Climate indices used for rule-based classification
Rule-based classification to partition homogeneous regions in terms of dominant flood generation processes.
Prediction of flood timing using the most relevant predictor for each hydro-climate class
Class 1
Class 2
Class 3
Class 4
Class 5
Hydro-climate classification system for flood timing prediction
MAP > 1200mm
YES
NO
Section 2.2 Observation and predictors of flood timing
Observed timing of annual maximum streamflow
Climatic predictors for flood timing
Section 2.1 Datasets
The Global Streamflow Indices and Metadata (GSIM) archive
ERA-Interim daily temperature and precipitation data
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Runoff DOYMAX
GLOBAL SCALE PREDICTION OF FLOOD TIMING

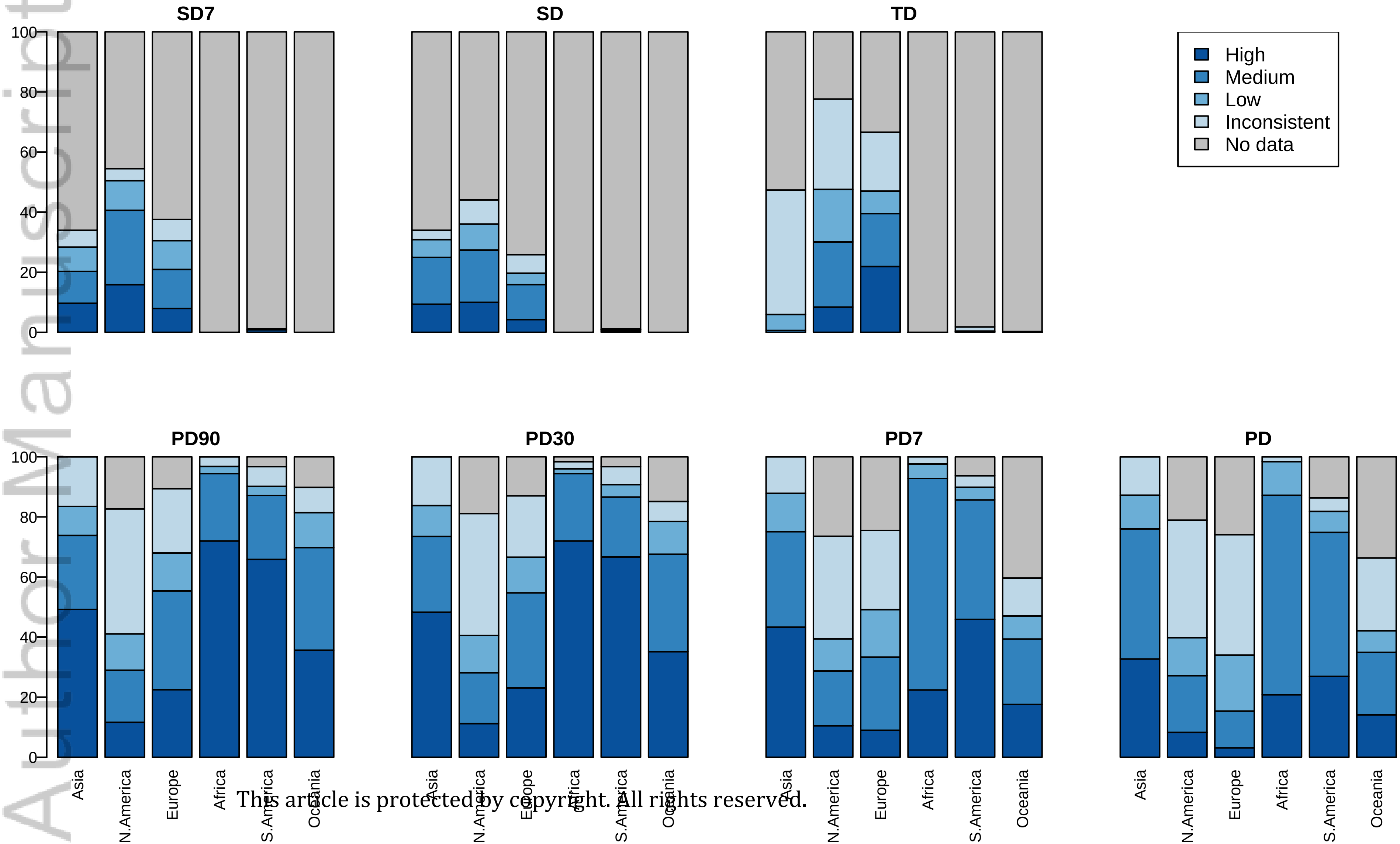

(a)
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
(b)
Dec
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov

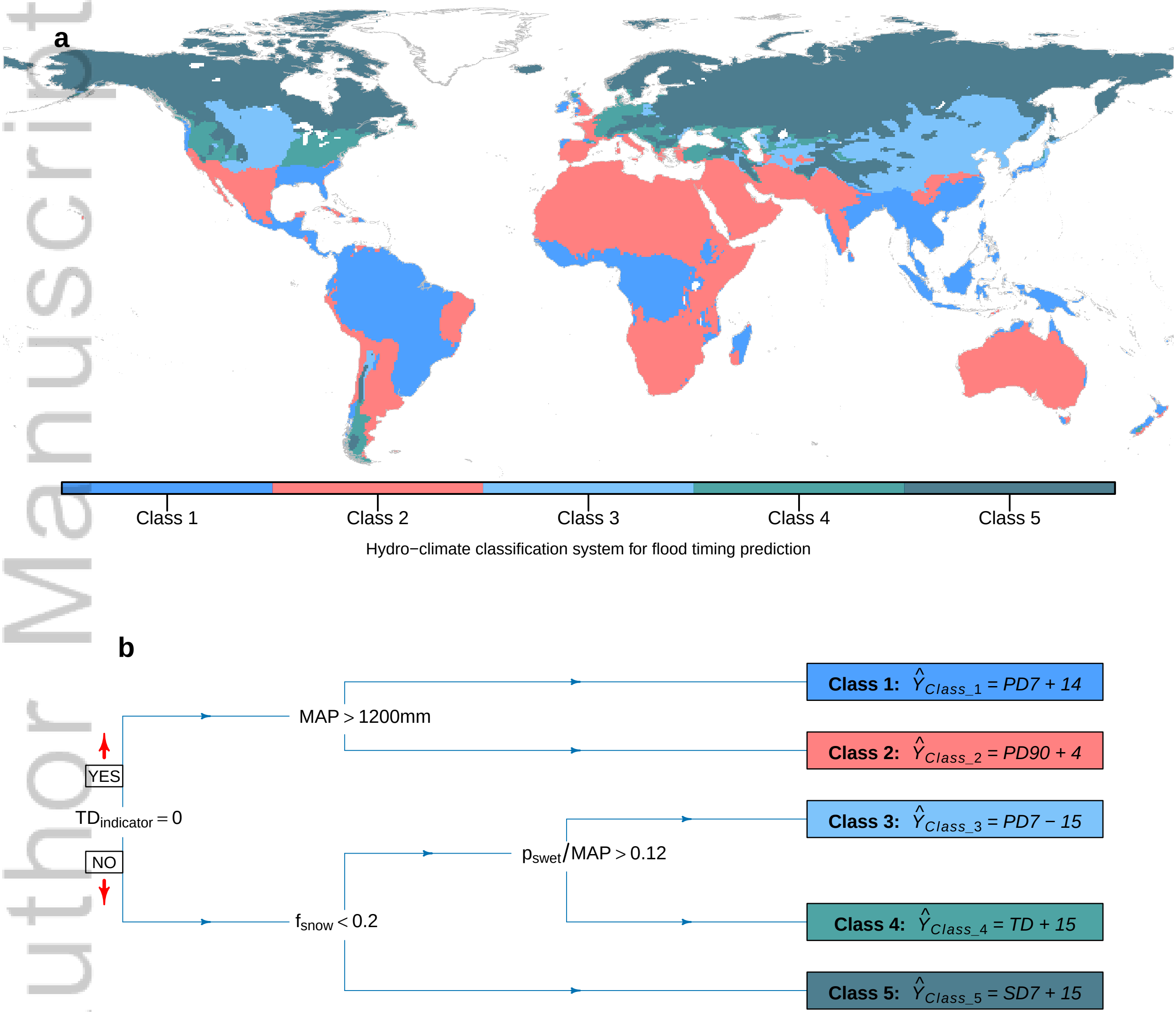

No predictor available
No rainfall−predictor available
No snowmelt−predictor available
Rainfall and snowmelt predictor available

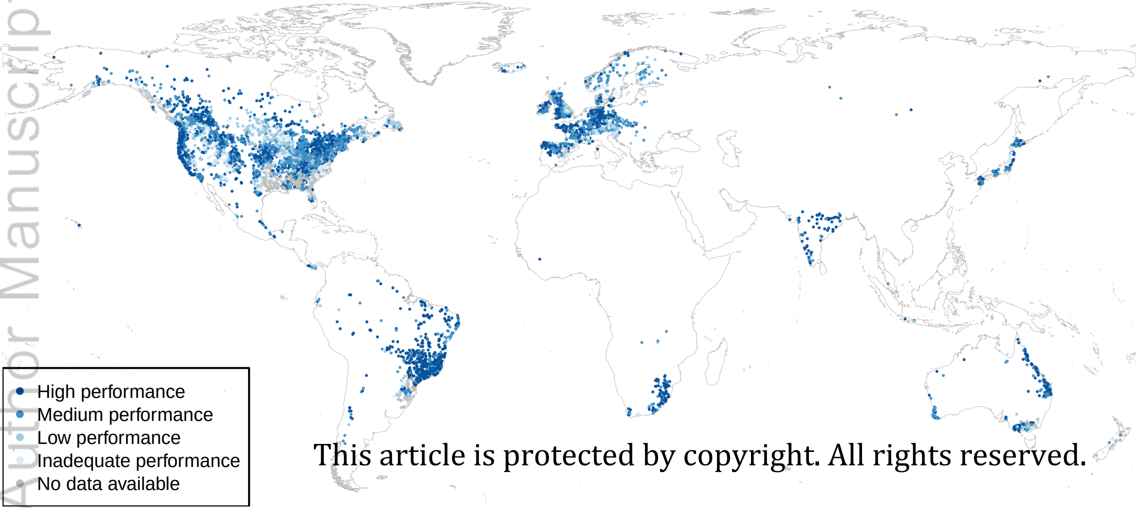

Previous non-terminal node
Non-terminal node n
Yes
$C_n > \alpha$
No
Terminal node R_j:
$\hat{Y}_{R_j} = X_{R_j} + \gamma_{R_j}$
Terminal node R_{j+1}:
$\hat{Y}_{R_{j+1}} = X_{R_{j+1}} + \gamma_{R_{j+1}}$

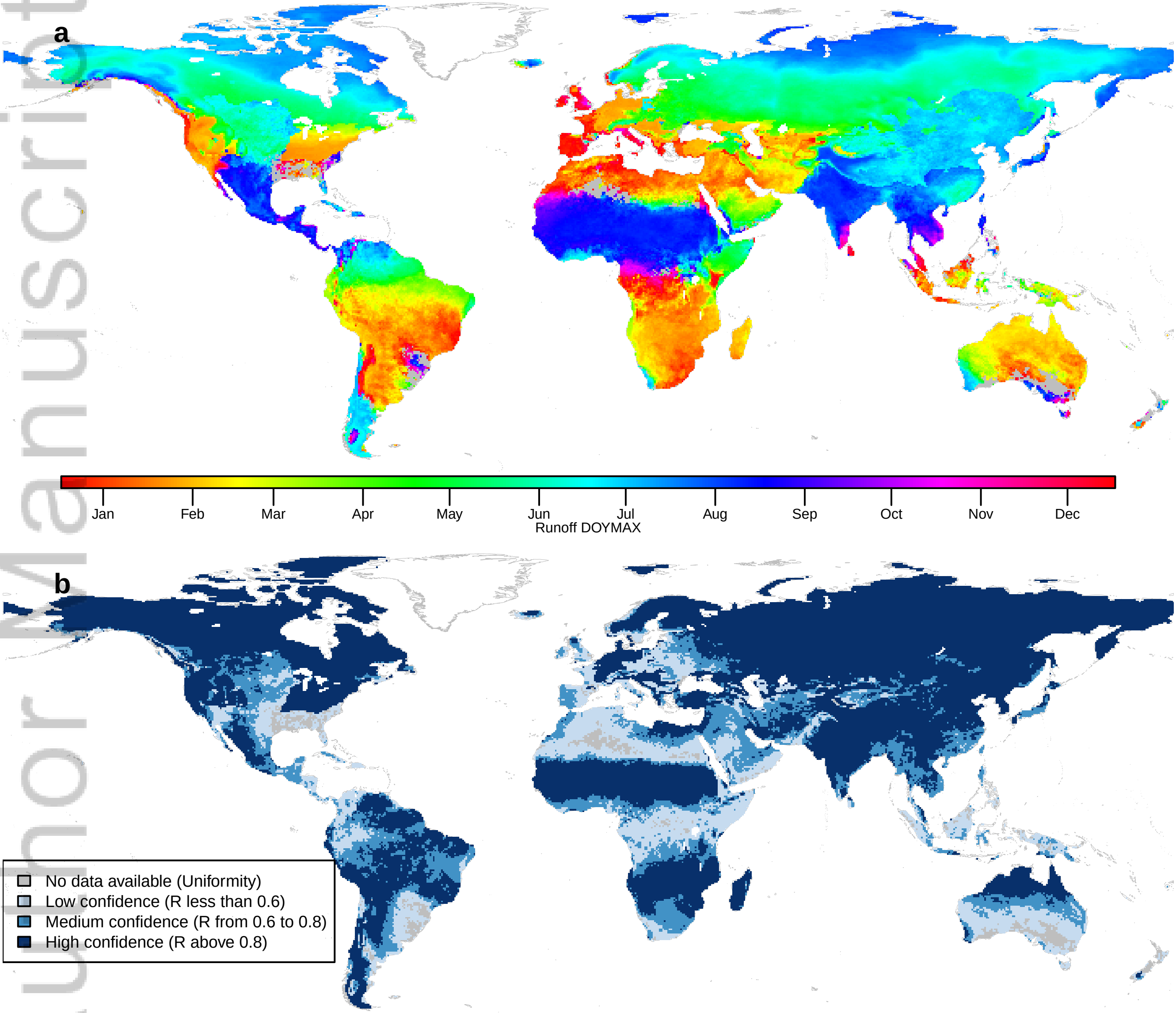

a
75
50
25
0
−25
−50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
b
75
50
25
0
−25
−50
R
Uniform
0−0.2
0.2−0.4
0.4−0.6
0.6−0.8
0.8−1


75
50
25
0
−25
−50
PD
PD7
PD30
PD90
TD
SD
SD7
no predictor


SD7
SD
TD
PD90
PD30
PD7
PD
High
Medium
Low
Inconsistent
No data
0
20
40
60
80
100
Asia
N.America
Europe
Africa
S.America
Oceania


a
Class 1
Class 2
Class 3
Class 4
Class 5
Hydro−climate classification system for flood timing prediction
b
MAP > 1200mm
YES
$TD_{indicator} = 0$
NO
$f_{snow} < 0.2$
$p_{swet}/MAP > 0.12$
Class 1: $\hat{Y}_{Class\_1} = PD7 + 14$
Class 2: $\hat{Y}_{Class\_2} = PD90 + 4$
Class 3: $\hat{Y}_{Class\_3} = PD7 - 15$
Class 4: $\hat{Y}_{Class\_4} = TD + 15$
Class 5: $\hat{Y}_{Class\_5} = SD7 + 15$


High performance
Medium performance
Low performance
Inadequate performance
No data available


a
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Runoff DOYMAX
b
No data available (Uniformity)
Low confidence (R less than 0.6)
Medium confidence (R from 0.6 to 0.8)
High confidence (R above 0.8)