2   *Hydrology and Earth System Sciences*

3   Supporting Information for

4   **Processes and controls of regional floods over eastern China**

5   Yixin Yang[1, 2], Long Yang[1, 2*], Jinghan Zhang[1, 2], Qiang Wang[1, 2]

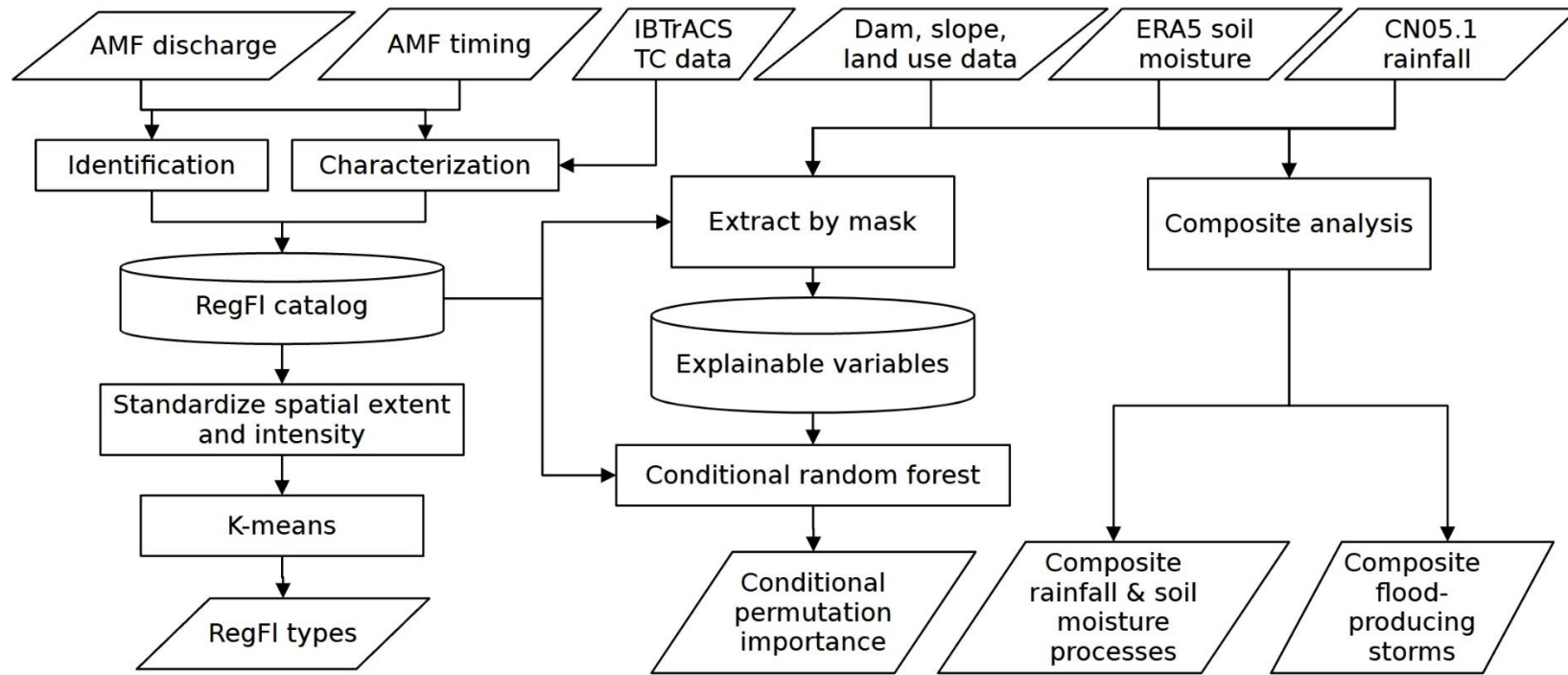
6   [1] School of Geography and Ocean Science, Nanjing University, Nanjing, China
7   [2] Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing, China

9
10   **Contents of this file**

11
12   Table S1
13   Fig. S1 to S7

14 Table S1. Key hyperparameters of conditional random forest model and out-of-bag (OOB) performance evaluation metrics for eight models.
15 Hyperparameter *mtry* is chosen from 2 to 11. Hyperparameter *ntree* is chosen from 50-500 trees with 50 trees as step. Other hyperparameters are
16 set as default. The optimal hyperparameter combination are selected to achieve the lowest RMSE. Non-parametric Mann-Whitney U test and
17 Kruskal-Wallis test show no significant difference between train error and OOB error, indicating little probability of overfitting. OOB error is the
18 prediction error (i.e., prediction minus true value) on each sample using only trees that did not contain this sample in their bootstrap sample.
19 Training error is the prediction error on each sample using training set.

Predictand	Key hyperparameters		Performance metrics (OOB)		P-value of difference test between train and OOB errors	
	mtry	ntree	RMSE	R-squared	Mann-Whitney test	Kruskal-Wallis test
RFI (All)	3	100	2.36	0.89	0.64	0.49
RFI (Mild)	4	100	1.01	0.86	0.70	0.49
RFI (Large)	4	350	3.64	0.66	0.85	0.48
RFI (Intense)	5	50	2.81	0.43	0.65	0.47
Spatial extent (All)	8	50	$5.36 \times 10^4 km^2$	0.93	0.49	0.49
Spatial extent (Mild)	11	500	$3.21 \times 10^4 km^2$	0.81	0.58	0.49
Spatial extent (Large)	7	50	$7.19 \times 10^4 km^2$	0.73	0.87	0.48
Spatial extent (Intense)	5	50	$5.36 \times 10^4 km^2$	0.53	0.74	0.47

20
21
22
23

24
25   Figure S1. Flowchart of regional flood analyses in this study.
26

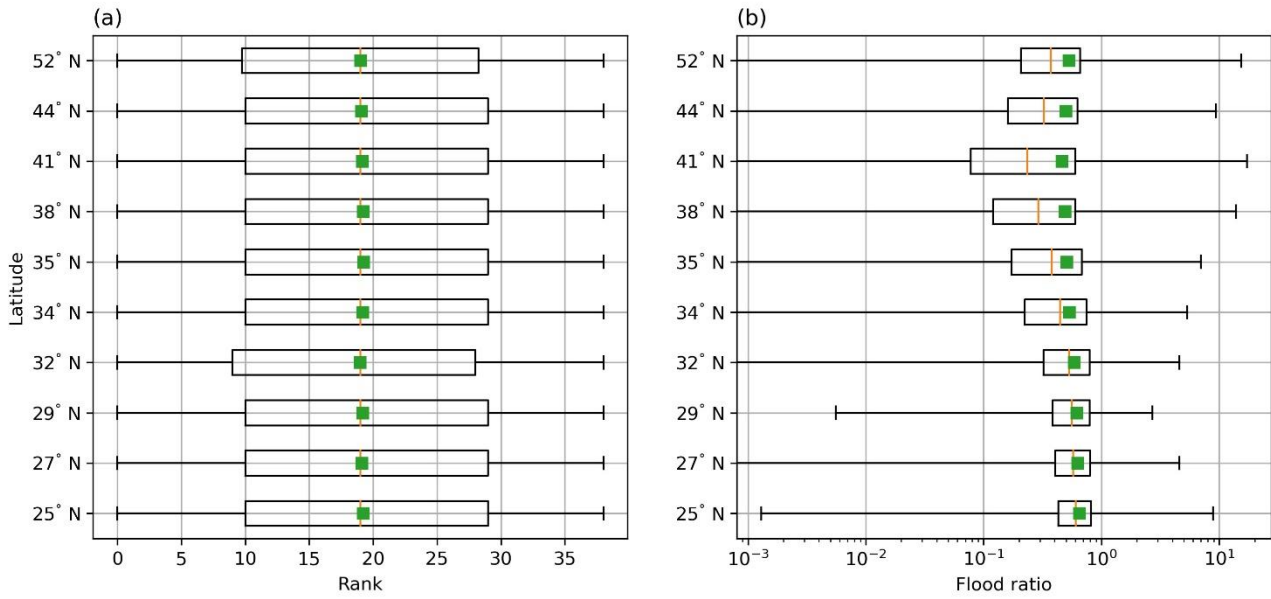



Figure S2. Decile boxplots for latitude versus (a) rank and (b) flood ratio (ratio of flood peak magnitude to the empirical 90th percentile of flood peaks). The orange line and green square within the box represent the median and mean values, respectively. The box spans the 25th and 75th percentile, and the whiskers represent the minimum and maximum values.

32

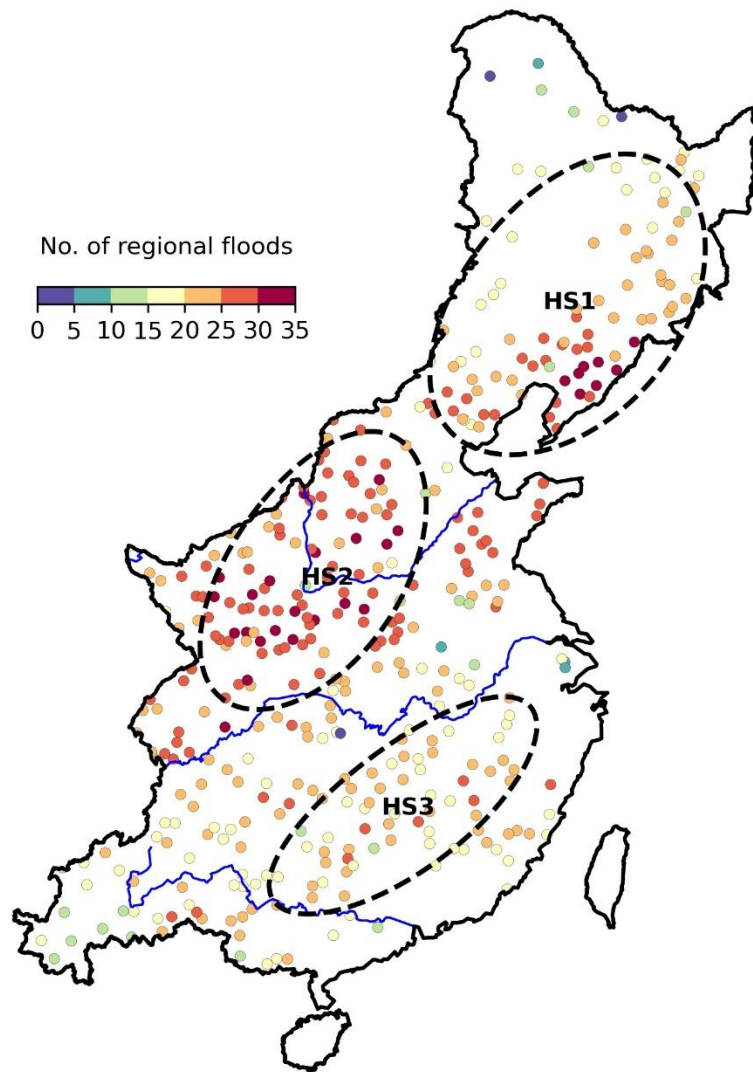
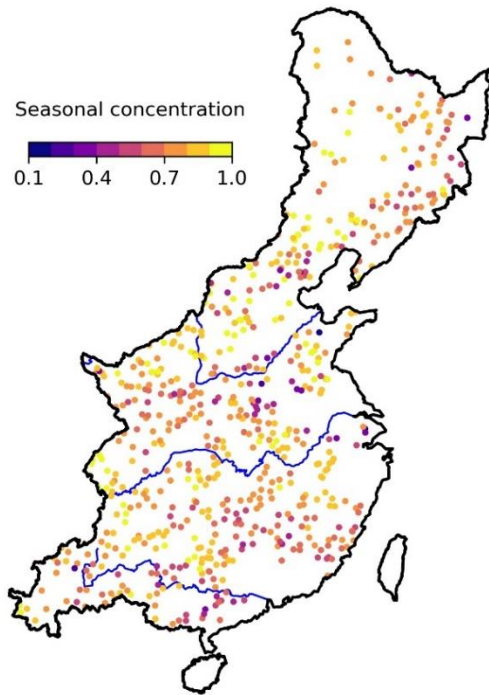
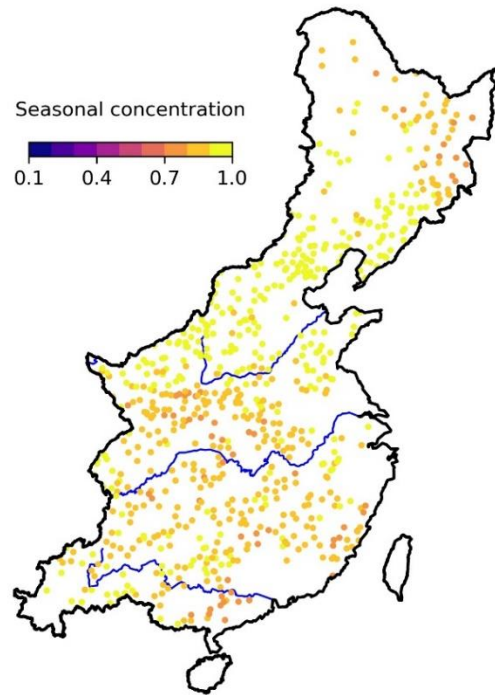



Figure S3. Spatial distribution of regional flood frequency using spatially uniform stations (i.e., non-
nested basins). The spatial pattern is similar to results obtained from all the stations (Fig. 3a).

35

36

37 Figure S4. Seasonal concentration of (a) isolated and (b) regional floods. Higher values indicate more
38 concentrated timing of AMFs.

39

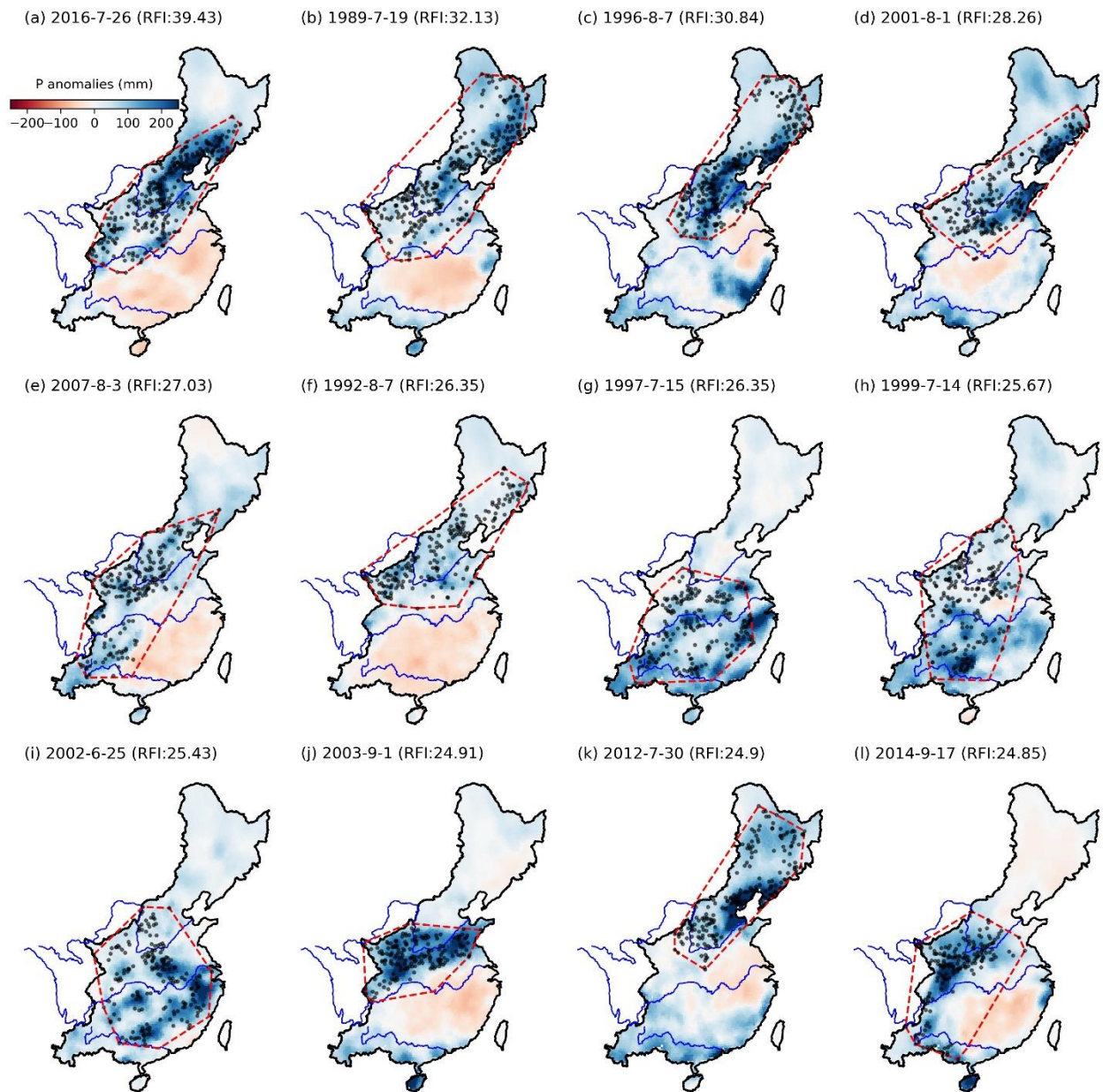



Figure S5. Precipitation anomalies for the top twelve severest *RegFls*. Black dots represent AMFs with drainage area smaller than 5,000 km^2 in the *RegFl*. Red dashed-line polygons represent convex hull.

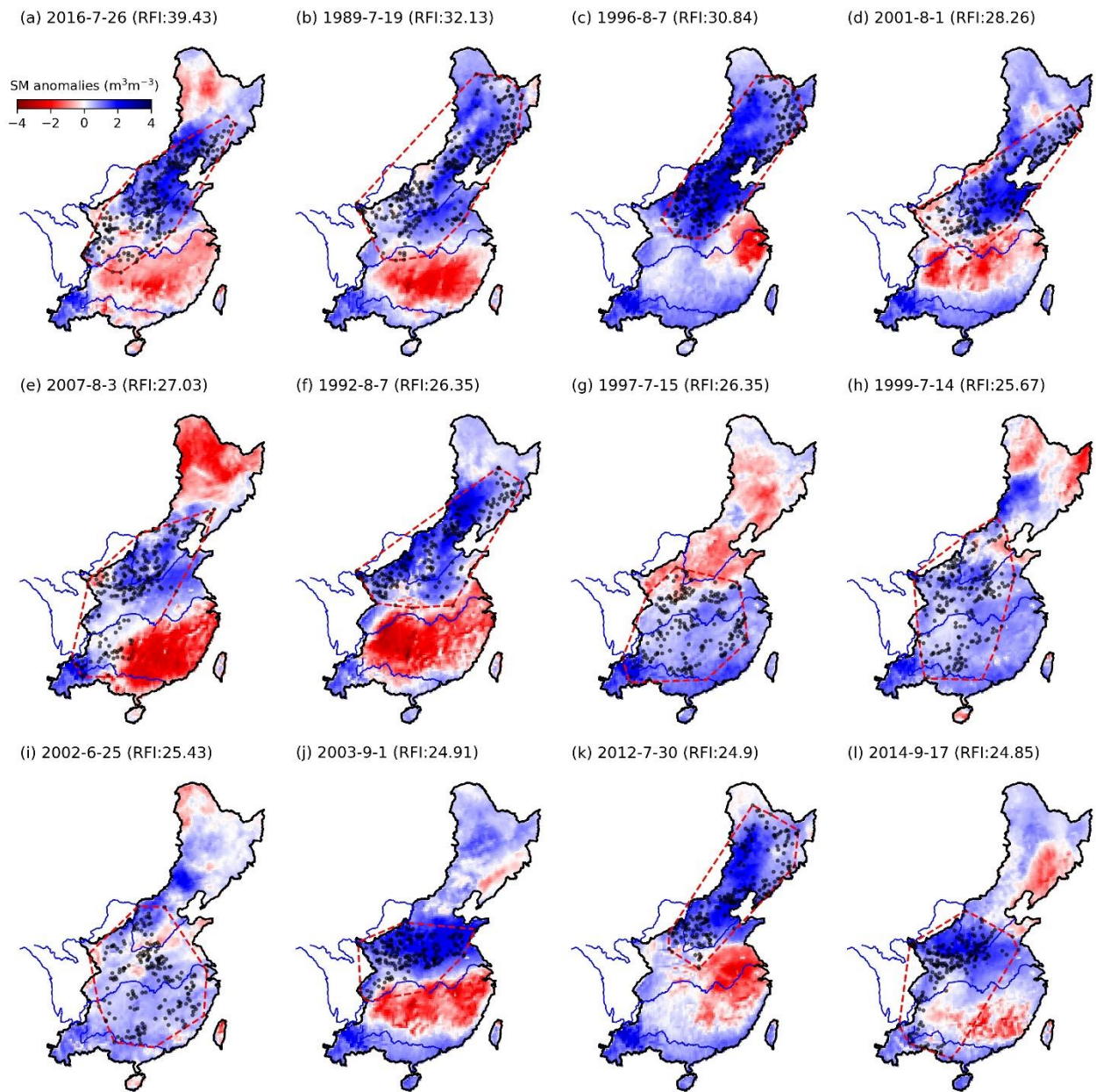



Figure S6. Soil moisture anomlies for the top twelve severest *RegFls*. Black dots represent AMFs with drainage area smaller than 5,000 km² in the *RegFl*. Red dashed-line polygons represent convex hull.

48

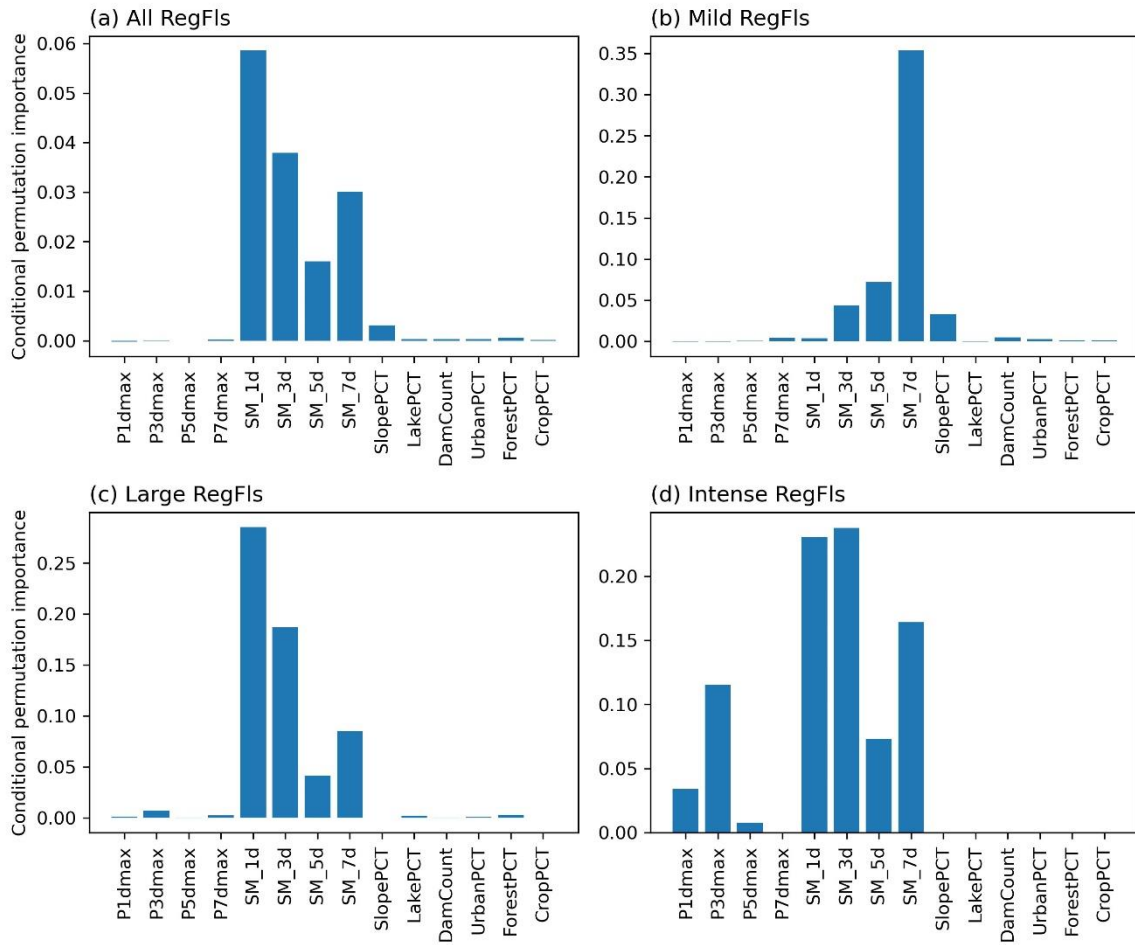



Figure S7. The conditional permutation importance for potential flood drivers in predicting *RegFls* spatial extent for (a) All, (b) Mild, (c) Large and (d) Intense *RegFls*. Refer to Table 1 for details of the variables.