Water Resources Research

Supporting Information for

Distributed Flashiness-Intensity-Duration-Frequency products over the conterminous US

Zhi Li[1,*], Shang Gao[2], Mengye Chen[1], Jiaqi Zhang[1], Jonathan J. Gourley[3], Humberto Veragra[4], Siyu Zhu[1], Sebastian Ferraro[1], Yixin Wen[5], Tiantian Yang[1], and Yang Hong[1*]

[1]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, USA

[2]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA

[3]NOAA/National Severe Storms Laboratory, Norman, OK, USA

[4]College of Engineering, University of Iowa, Iowa City, IA, USA

[5]Department of Geography, University of Florida, Gainesville, FL, USA.

Contents of this file

Additional Supporting Information (Files uploaded separately)

None

Introduction

This supplementary file contains Figures (Figs. S1 and S2) showing explanatory results interpreted from machine learning model to enrich our discussion, and Tables that detail

basin attributes used in this study for model training (Table S1) and description of all machine learning models (Table S2).
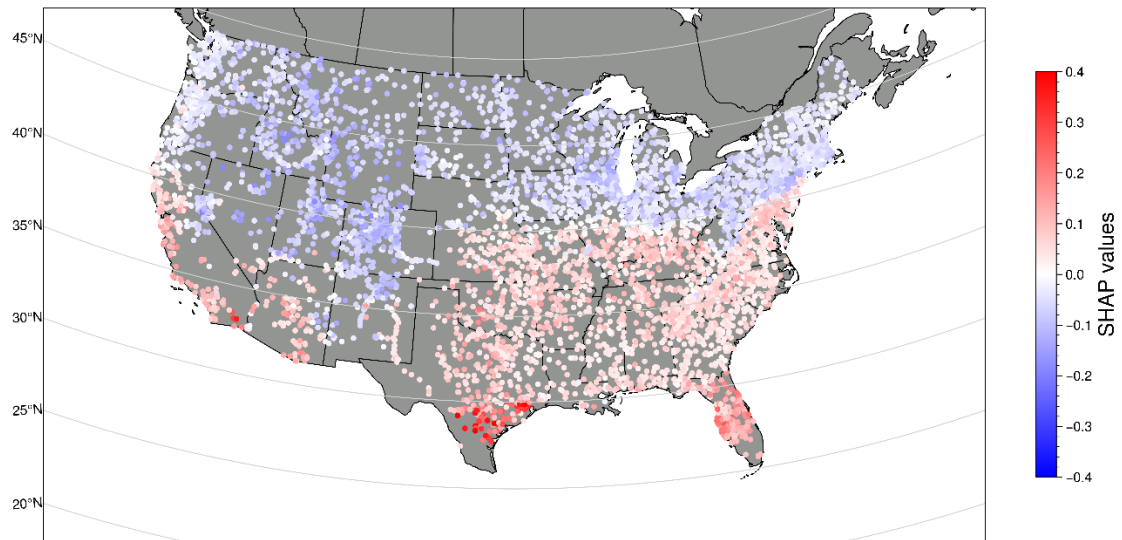

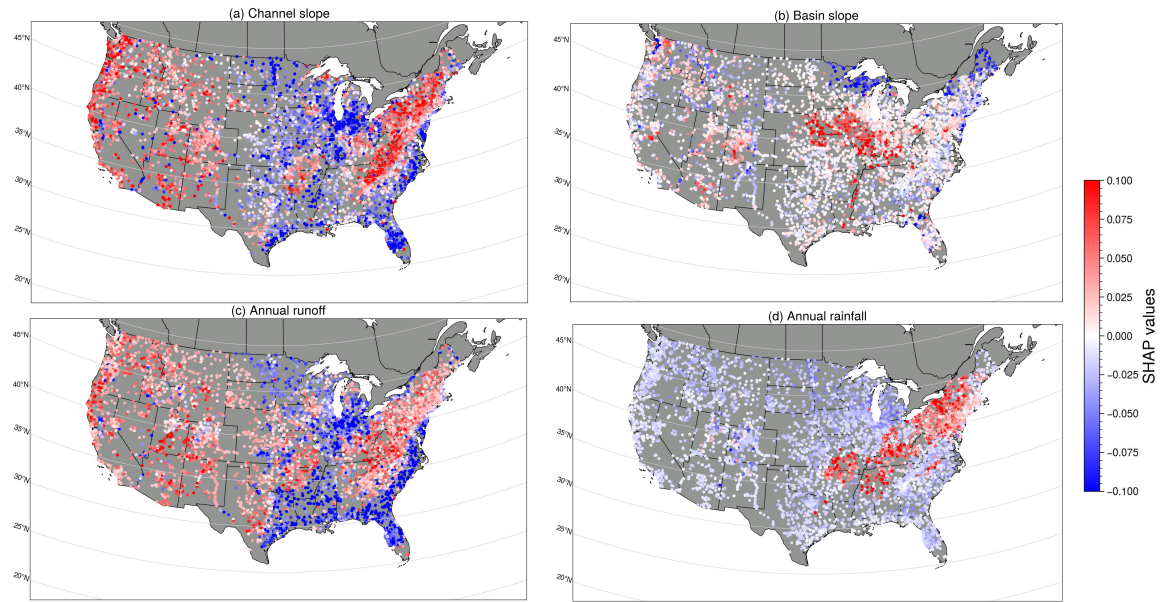**Figure S1.** Map of the SHAP values for air temperature variable.

Figure S2. Similar to Figure 1, but for (a) channel slope, (b) basin slope, (c) Annual runoff, and (d) Annual rainfall.

Table S1. A table of all basin attributes used for model training

Class	Attribute name	Unit	Source data
Hydrology	Natural discharge	m^3/s	WaterGAP v2.2
	Surface runoff	mm	WaterGAP v2.2
	Inundation extent	%	GIEMS-D15
	Groundwater table depth	cm	Fan et al. (2013)
	Degree of regulation	%	GRanD
	River area	Hectares	HydriSHEDS and WaterGAP v2.2
	River volume	1000 m^3	HydriSHEDS and WaterGAP v2.2
Physiography	Channel slope	dm/km	EarthEnv-DEM90
	Basin slope	deg	EarthEnv-DEM90
	Elevation	m	EarthEnv-DEM90
	Drainage area	km^2	EF5
Climate	Annual precipitation	mm/year	WorldClim v1.4
	Actual evaporation	mm/year	Global High-Resolution Soil-Water Balance
	Average annual air temperature		WorldClim v1.4 and Global-PET v1
	Potential evaporation	mm/year	Global-PET v1
	Snow cover extent	%	MODIS/Aqua Snow Cover
	Moisture index	-	WorldClim v1.4 and Global-PET v1

	Aridity index	-	Global Aridity Index v1
Soils & Geology	Soil water content	%	Global High-Resolution Soil-Water Balance
	Clay fraction	%	SoilGrids1km
	Silt fraction	%	SoilGrids1km
	Sand fraction	%	SoilGrids1km
	Karst area extent	%	World Map of Carbonate Rock Outcrops v3.0
	Soil erosion	kg/hectare/year	RUSLE-based Global Soil Erosion Modelling platform v1.2
Anthropogenic	Road density	m/km^2	Global Roads Inventory Project v4
	Urban extent	%	Global Human Settlement (GHS) Settlement Model v1.0
	Population density	People/km^2	Gridded Population of the World (GPW) v4
Landcover	Tree (broadleaved/evergreen)	%	GLC2000
	Tree (broadleaved/deciduous)	%	GLC2000
	Tree (needle-leavd/evergreen)	%	GLC2000
	Tree (mixed leaf)	%	GLC2000
	Tree (mosaic)	%	GLC2000

	Shrub (evergreen)	%	GLC2000
	Shrub (deciduous)	%	GLC2000
	Herbaceous	%	GLC2000
	Sparse herbaceous	%	GLC2000
	Flooded shrub	%	GLC2000
	Cultivated land	%	GLC2000
	Mosaic	%	GLC2000
	Water body	%	GLC2000
	Snow	%	GLC2000
	Artificial	%	GLC2000
Natural vegetation	Tropical evergreen	%	EarthStat
	Tropical deciduous	%	EarthStat
	Temperate deciduous	%	EarthStat
	Temperate evergreen	%	EarthStat
	Boreal evergreen	%	EarthStat
	Boreal deciduous	%	EarthStat
	Evergreen	%	EarthStat
	Savanna	%	EarthStat
	Grassland	%	EarthStat
	Dense shrub	%	EarthStat
	Open shrub	%	EarthStat

	Tundra	%	EarthStat
	Desert	%	EarthStat
Wetland	Lake	%	Global Lakes and Wetlands Database
	Reservoir	%	Global Lakes and Wetlands Database
	River	%	Global Lakes and Wetlands Database
	Peatland	%	Global Lakes and Wetlands Database

Table S2 Description of Machine Learning models used in this research

Model	Description	Type
Linear Regression	Linear model $y = w_i x_i + b_i$ solves for the weights w_i and b_i by minimizing the residual sum of squares.	Linear
Lasso regression	Linear regression with L1 regression. It adds a penalty equal to the absolute value of the magnitude of coefficients to the loss function to avoid overfitting.	Linear
Ridge Regression	Linear regression with L2 regression. It adds a penalty equal to the square of the magnitude of coefficients to the loss function to avoid overfitting	Linear
Bayesian ridge	It is a form of ridge regression where the prior over the coefficients is a Gaussian distribution	Linear
Elastic Net	It Combines L1 and L2 regularization of Lasso and Ridge methods.	Linear
Lasso Least Angle Regression	It Uses a type of linear regression that uses L1 regularization to estimate the sparse solutions, i.e., solutions that have fewer non-zero coefficients.	Linear
Least angle regression	A regression algorithm for high-dimensional data. LARS is similar to forward stepwise regression but less greedy	Linear
Huber regressor	Linear regression model that is robust to outliers.	Linear
Orthogonal Matching Pursuit	It is a sparse approximation algorithm to find the "best" matching of multidimensional data onto the span of an over-complete dictionary.	Linear
Dummy regressor	A regressor that makes predictions using simple rules. It is a baseline to compare to other regressors.	Linear
Passive Aggressive Regressor	An online learning algorithm. It remains passive for a correct classification outcome and turns aggressive in the event of a miscalculation, updating, and adjusting	Linear
Decision trees	A model makes its predictions by learning simple decision rules inferred from the data features.	Tree-based
K Neighbors Regressor	An instance-based learning or non-generalizing learning. It does not attempt to construct a general internal model, but simply stores instances of the training data.	Instance-based
Random Forests	An ensemble learning method that operates by constructing multiple decision trees at training time and outputting the mean prediction of the individual trees	Ensemble-based
Gradient Boosting Regressor	Builds an additive model in a forward stage-wise fashion. It allows for the optimization of arbitrary differentiable loss functions	Enseble-based

Light Gradient Boosting Machine	A gradient boosting framework that uses tree-based learning algorithms. It is designed to be distributed and efficient with faster training speed and higher efficiency	Ensemble-based
Extreme Gradient Boosting (XGBoost)	An optimized distributed gradient boosting method designed to be highly efficient, flexible, and portable.	Ensemble-based
Extra trees regressor	It implements a meta estimator that fits several randomized decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting	Ensemble-based
AdaBoost regressor	An ensemble learning method that constructs a classifier by fitting the original dataset and then fits additional copies of the classifier on the same dataset but adjusts the weights of incorrectly classified instances	Ensemble-based