Water Resources Management (2019) 33:3449–3468
https://doi.org/10.1007/s11269-019-02310-y

Disentangling the Contributions of Climate and Basin Characteristics to Water Yield Across Spatial and Temporal Scales in the Yangtze River Basin: A Combined Hydrological Model and Boosted Regression Approach

Fengyun Sun[1,2,3] • Alfonso Mejia[4] • Yue Che[1,2,3]

Received: 31 October 2018 / Accepted: 20 June 2019 /
Published online: 23 July 2019

Abstract
The dependence and contribution of explanatory variables or predictors to water yield need to be closely analyzed and accurately quantified to better understand water balances as well as for effective water resources management. It is generally challenging, however, to disentangle the contribution of individual climate variables from that of basin characteristics to the integrated water yield response. Here we propose a method to concurrently quantify and analyze the effects of climate and basin predictors on water yields. This method employs the Soil and Water Assessment Tool (SWAT) to simulate water yield. Simulated results are then analyzed and compared using Boosted Regression Trees (BRTs) at multiple spatial and temporal scales. Results indicate that in the Yangtze River Basin (YRB) on average, precipitation is of paramount importance, followed by land cover, while slope has the lowest contribution. The average relative contributions of soil moisture, maximum and minimum temperatures are different among temporal scales. More stable and reliable results are derived at the daily scale compared to the yearly and monthly scale. Our results make evident that generalizations about water yield response made in the absence of a comprehensive and accurate description of site- and scale-specific contributions can lead to misleading assessments. This proposed approach can be useful for informing and supporting more effective water resources management goals.

Keywords Water yield · Spatial and temporal scales · Soil and water assessment tool (SWAT) · Boosted regression tree (BRT) · Yangtze River Basin

✉ Yue Che
yche@des.ecnu.edu.cn

Extended author information available on the last page of the article

1 Introduction

It has long been recognized that climate conditions and basin physical characteristics are dynamically coupled through the physical processes that transport and store water across the land surface and subsurface. Water yield, defined as the total amount of water leaving a hydrological response unit and entering the stream network during a given period (Arnold 2012; Stone et al. 2003), is of great importance as it supplies water resources to human being and natural resources. However, water yield has a highly non-linear response to and complex interaction with major environmental factors such as climate conditions and basin physical characteristics (Feng et al. 2012; Liu et al. 2011), resulting in highly variable and contradictory results in previous studies (Berg et al. 2016; Pessacg et al. 2015; Rice et al. 2015; Sun et al. 2006; Zhou et al. 2015). There is a need to explore the effect of multiple factors, both meteorological and landscape-related, on water yield (Feng et al. 2012; Pessacg et al. 2015).

Case studies of paired watershed experiments worldwide have reported the effects of climate and forestry variability on water yield, e.g., Brown et al. (2005), Gokbulak et al. (2016), Wu et al. (2015), and Yao et al. (2015). Based on these evidence, studies have sought to generalize our understanding of water yield variations. For example, Zhou et al. (2015) proposed a unifying theory to describe the relative influence of climate and basin factors on water yield response globally based on the relationship among water yield, a wetness index, and basin characteristics, deepened our understanding of the dependence of global hydrological response patterns to land cover and climate; Wang et al. (2011) obtained a statistical relationship between water yield and some of its predictors (e.g., precipitation, forest cover, and altitude) in northern China. Despite these efforts, describing the relative roles played by different climate variable and basin characteristic in determining water yield is largely absent. It can serve to improve our understanding of the main underlying mechanisms governing basins, in support of a variety of water resources management goals.

Progress has been made during the past decades in understanding how water yield varies with scale (Berg et al. 2016; de Vente and Poesen 2005; Kuria and Vogel 2015; Wang et al. 2010). There is now strong experimental evidence indicating that scale is a key control variable for water yield studies, and perhaps one of the main reasons for why past studies are often inconclusive or show conflicting results (McDonnell et al. 2007). The influence of spatial scales on water yield has been quantified and detected at the site (Lu et al. 2013), basin (Gharun et al. 2014; van Dijk et al. 2012; Wang et al. 2011), country (Sun et al. 2006), and global (Farley et al. 2005) scale. The basin scale ranging from small, ~148 km^2 (Gharun et al. 2014), to very large, ~1,300,000 km^2 in the case of the Missouri River basin, United States (Stone et al. 2001). The influence of temporal scales on water yield has been quantified primary at the yearly (Adams and Fowler 2006; Yao et al. 2015) or seasonal (monthly) scale (Feng et al. 2012), while studies that consider the daily scale are notably rare, particularly for large basins. The effects of predictors on monthly and daily water yield are less well understood, even though the effects can be as or more important than at the yearly scale (Brown et al. 2005), depending on the application. Indeed, it is currently difficult to understand and assess how water yield changes with the temporal scale since studies tend to consider a single scale. To the best of our knowledge, no systematic study about the effects of predictors on water yield across both spatial and temporal scales has been undertaken.

In aiming to close the aforementioned knowledge gaps in describing the relative roles of different climate variables and basin characteristics in determining water yield across both the spatial and temporal scales. Our overall objectives with this study are i) to disentangle the relative contributions of climate and basin characteristics to total water yield for the Yangtze River Basin; ii) to identify the marginal effects of predictor variables on water yield; and iii) to detect the influencing effects of different spatial and temporal scales on relative contributions of predictors to water yield.

2 Methods and Data

2.1 Study Area

As the study area, the Yangtze River Basin (YRB) is used. The YRB, with a drainage area of ~1.8 million km^2 and spatial extent of 24°30′ - 35°45′ N and 90°33′ - 112°25′ E, is the third largest river basin on Earth. It originates in the Tibetan Plateau, crossing China from west to the east and finally flowing into the East China Sea (Fig. 1), draining nearly 20% of the land area of the country (Xu et al. 2008). Characterized by great disparity of terrain, the YRB is comprised by a diverse physiography, climate, hydrology, soil, and other natural geographic conditions. The long-term annual precipitation ranges from 500 mm in the west to 2,500 mm in the east, with an average of 1,070 mm. More than 60% of the annual precipitation is brought by the monsoon winds and falls primarily as rain in the summer months of June – August (Zhang et al. 2008). The Yangtze carries more water than any other river in China.

2.2 Water Yield Estimation

Water yield is obtained in this study via simulation using the Soil and Water Assessment Tool (SWAT). Compared with using statistical relationships between water yield and selected

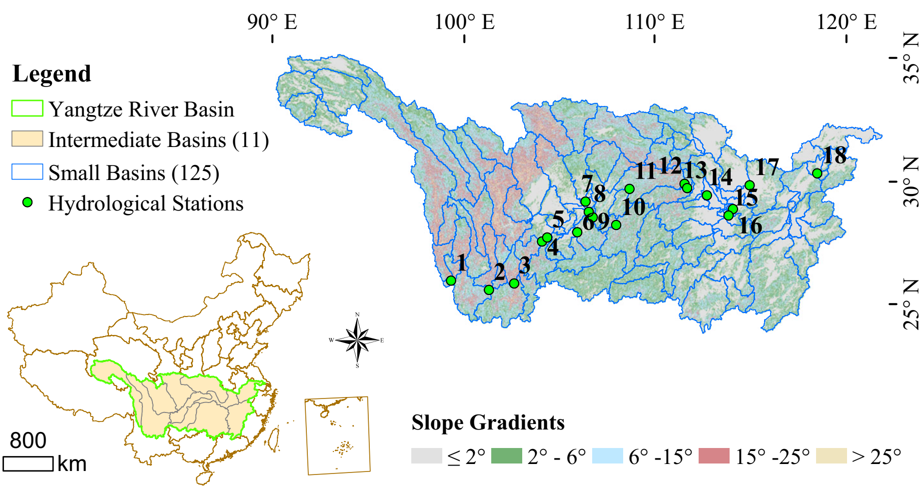

Fig. 1 Map illustrating the location, basin boundary, hydrological stations and slope gradients of the Yangtze River Basin

variables, it is preferred, whenever feasible, to directly model water yield by taking into consideration the physical processes involved (Neitsch et al. 2009). The SWAT divides the river basin into a number of hydrological response units (HRUs), each of them possessing a unique land cover/soil attribute/slope gradient identifier. The water yield for the basin is predicted as the total amount of water leaving the HRUs and entering the stream network during a time step (Arnold 2012; Neitsch et al. 2009). The basis of the water yield simulation is given by

$$\text{Wateryield} = \text{SR} + \text{LF} + \text{GD–TL–PA},$$

where SR, LF, GD, TL and PA, respectively, stand for surface runoff, lateral flow, groundwater discharge, transmission losses and pond abstractions.

In this study, the YRB is divided into 20,215 HRUs. To assess the effect of spatial scale on water yield and its predictors, the 20,215 HRUs are analyzed at three different spatial scales: i) the large basin scale, i.e., the whole YRB with a drainage area of 1,726,335.81 km^2; ii) the intermediate basin scale, i.e., the YRB divided into 11 adjoining intermediate basins with drainage areas ranging from 83,651.28 km^2 to 264,571.96 km^2, with the median being 164,564.14 km^2; and iii) the small basin scale, i.e., the YRB divided into 125 adjoining small basins with drainage areas ranging from 6.12 km^2 to 85,723.55 km^2, with the median being 11,190.58 km^2. Figure 1 illustrates the small and intermediate basin scale. In order to account for different temporal scales, the simulations are conducted at the yearly, monthly and daily scales using ten years of data (1994–2013). The SWAT model was built, monthly calibrated and implemented by Sun et al. (2016). Herein, we further calibrate the SWAT model at daily scale to increase the modeling accuracy. A consecutive 21-year long daily discharge observations at 18 hydrological stations are used for the calibration. The calibration and validation period are 15 years (Jan. 1st 1990 - Dec. 31st 2004) and 6 years (Jan. 1st 2005 - Dec. 31st 2010), respectively. The performance of the SWAT model is routinely measured by the Coefficient of determination (R^2), which measures the proportion of the variance in measured data explained by the modeled data and the Nash-Sutcliffe coefficient (NS), which measures the relative magnitude of the residual variance compared to the measured data variance. R^2 ranges between 0 and 1, the higher the value the better the fit, and values greater than 0.5 are considered acceptable. NS ranges from $-\infty$ to 1, with the performance rating acceptable if values between 0 and 1 (Castillo et al. 2014). Overall, with a medium (range) R^2 and NS equal to 0.84 (0.44–0.9) and 0.71 (0.32–0.89) (Table 1 and Fig. 2), respectively, for the combined calibration and validation period (1990–2010), the established SWAT model quite accurately reproduces the rainfall–runoff process for the YRB, and is suitable to investigate the water balance processes in the YRB.

To understand the importance of different components in predicting water yield, the analysis is herein focused on a few major environmental variables, including precipitation, maximum and minimum temperature, land cover, soil, and slope (Table 2). Note that the land cover, soil and slope variables are comprised by 9, 139, and 5 different categories, respectively. We obtained these categories from various datasets as described by Sun et al. (2016). These variables are key providers and mediators of moisture and energy which drive water yield processes and runoff potential. These variables have been identified and used in several previous water yield studies (Mehta et al. 2011; Wang et al. 2010). They have been proven to be useful and relevant for understanding and modeling water yield.

Table 1 Summary of the SWAT model performance

Streamflow gauge name	Gauge No.	Calibration		Validation	
		R^2	NS	R^2	NS
Shigu	1	0.71	0.32	0.84	0.68
Panzhihua	2	0.80	0.67	0.85	0.74
Huatan	3	0.88	0.67	0.90	0.68
Pingshan	4	0.88	0.70	0.90	0.70
Gaochang	5	0.61	0.47	0.62	0.43
Zhutuo	6	0.83	0.72	0.87	0.72
Wusheng	7	0.44	0.41	0.55	0.55
Beibei	8	0.49	0.48	0.53	0.53
Cuntan	9	0.83	0.77	0.82	0.76
Wulong	10	0.52	0.46	0.57	0.40
Wanxian	11	0.84	0.78	0.84	0.79
Huanglingmiao	12	0.85	0.84	0.79	0.74
Yichang	13	0.86	0.81	0.79	0.74
Shashi	14	0.87	0.83	0.81	0.70
Chenglingji	15	0.63	0.61	0.62	0.60
Luoshan	16	0.89	0.87	0.87	0.80
Hankou	17	0.90	0.89	0.88	0.86
Datong	18	0.86	0.83	0.86	0.83

2.3 Statistical Analysis by BRT Model

It is generally challenging to separate the contribution of individual variables to the integrated water yield response, while the Boosted Regression Tree (BRT) models is skilled in analyzing the strong non-linearity between water yield response and major environmental variables. BRTs estimate the importance of each predictor through a hierarchical binary splitting procedure that maximizes differences among variables. At each step, a forward stagewise process reserves existing trees that already built and adds new trees by reweighing residuals from previous trees. The BRT models therefore produce thousands of trees and a mean variable importance estimate is extracted across all trees (De'ath 2007; Elith et al. 2008; Friedman and Meulman 2003; James et al. 2013). The procedure is summarized in Fig. 3, $f(x)$ approximates the relationship between response variable y and predictor variables x. The βm represent weights given to the nodes of each tree and determine how predictions from each of the trees are combined. The parameters β_m and γ_m are estimated by minimizing a specified loss function sequentially from 1 to M (De'ath 2007).

This BRTs also produce partial dependence plots to assist with visualizing the “average” effect of each predictor variable on the response variable (Golden et al. 2016). The superior performance of BRTs in dealing with problems requiring the separation of complex and highly interdependent variables has been widely recognized in recent years (Liu et al. 2016; Price 2011). Despite several applications on hydrology related topics, e.g., Hale et al. (2014); Golden et al. (2016); Salazar et al. (2016); Naghibi et al. (2018), the BRT models are still rarely used.

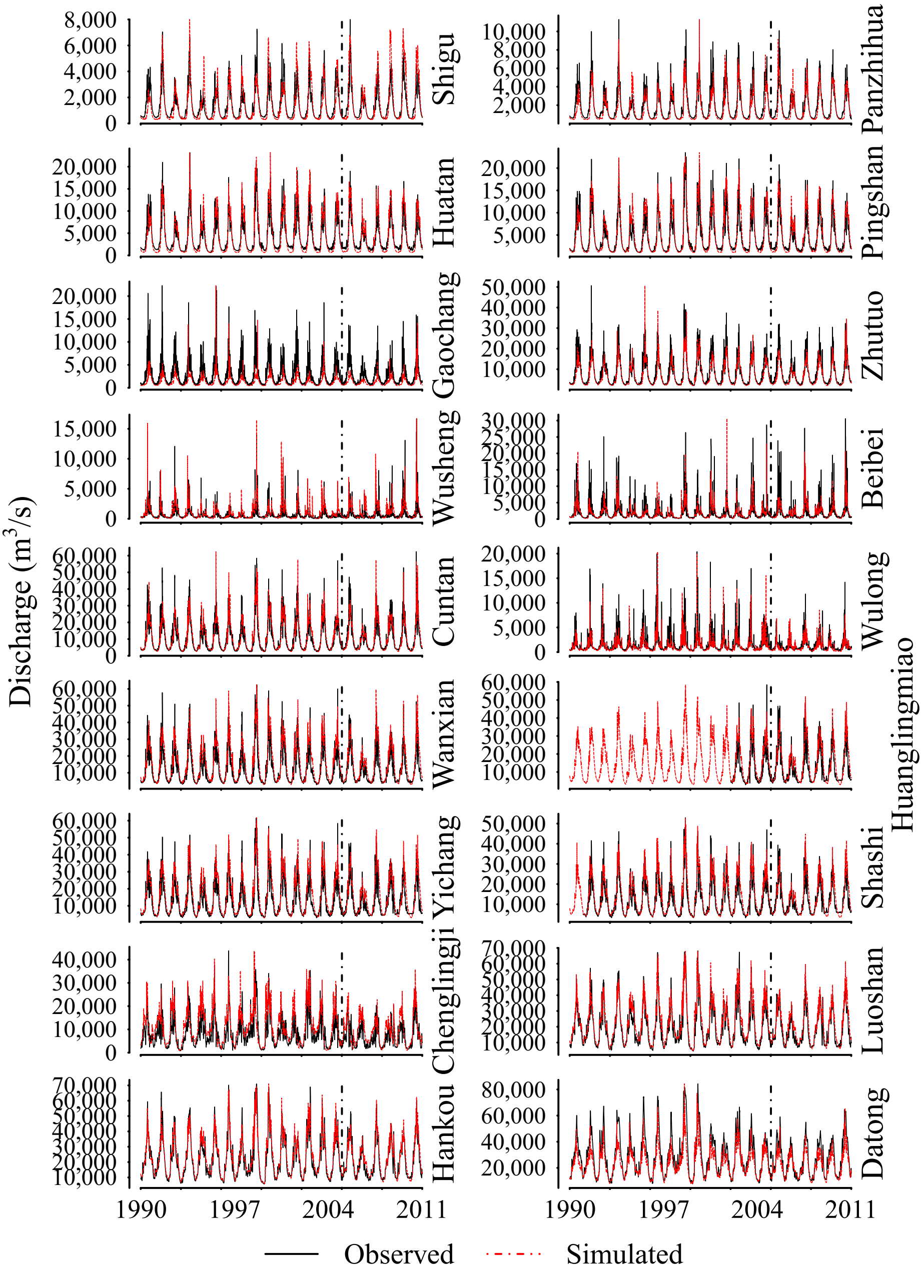

Fig. 2 Comparison of the observed and simulated streamflow for the calibration and validation period

For this study, 411 BRT models are used (375 models for the small basin scale, 33 for the intermediate basin scale and 3 for the large basin scale) to analyze the effect of spatial scale on water yield and its predictors. The BRT models are developed by means of the R-software (R Core Team 2014), specifically by using its package 'dismo' (Elith and

Table 2 Summary of candidate predictor variables used to develop the BRT model of water yield

Variable	Definition	Types (unit)	Mean ± SD	Range
Water yield	Dependent variable	Continuous (mm)	1.6 ± 4.0	0 – 405.9
Land cover	Independent variable (9 categories)	Nominal (unitless)	/	/
Soil	Independent variable (139 categories)	Nominal (unitless)	/	/
Slope	Independent variable (5 categories)	Ordinal (unitless)	/	/
Precip	Independent variable	Continuous (mm)	3.0 ± 10.2	0 – 423.8
MaxTem	Independent variable	Continuous (°)	21.2 ± 9.7	−15.5 – 42.8
MinTem	Independent variable	Continuous (°)	11.5 ± 10.0	−29.5 – 32.6

Precip Precipitation, *MaxTem* Maximum temperature, *MinTem* Minimum temperature

Leathwick 2015) with 'gbm' functions (Ridgeway 2007). According to the recommendations by Elith et al. (2008), all the BRT models used in this study have a learning rate of 0.005, a bag fraction of 0.5, a tree complexity of 5, and 10-fold cross validation. The performance obtained for the BRT models is adequate for assessing the relative contribution of climate and basin characteristics to determining water yield. The comparison of the average BRT results is validated via the one-way ANOVAS followed by Tukey's honest significant difference (HSD) test, where conditions of normality and homogeneity of variance are not met, non-parametric Krukal-Wallis tests are used.

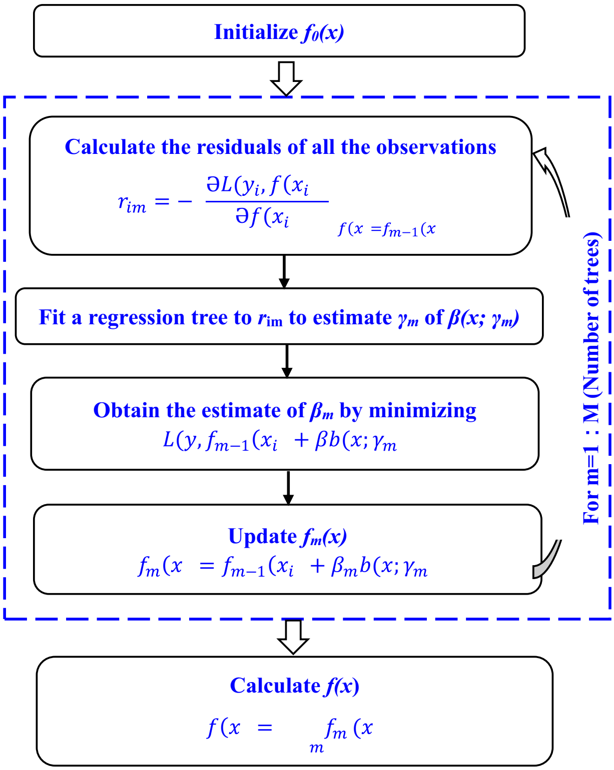

Fig. 3 BRT algorithm

3 Results

3.1 Relative Contribution of the Predictor Variables to the Total Water Yield

In this subsection, based on the 411 fitted BRT models, the results of the relative contribution of the six predictor variables to the total water yield are summarized and presented. The results are presented for the different basin (i.e., small, intermediate and large) and temporal (i.e., yearly, monthly and daily) scales considered. At the small basin scale, the relative contribution of the individual variables to the total water yield exhibits remarkable differences. The BRT analysis results at the small basin scale are illustrated and summarized in Fig. 4 and Table 3. For the small basin scale, the ranking of the mean contribution of each of the predictors to the total water yield indicates that precipitation is the most important predictor across the three temporal scales, followed by land cover, whereas slope is the least important having the lowest contribution to water yield ($P < 0.001$) (Table 3). The average contributions of the predictors based on the yearly and daily simulations have the exact same ranking orders; whereas in the case of the monthly simulations, the relative contribution of the minimum and maximum temperature outnumber the soil contribution to the water yield (Table 3).

Although the ranking values of the six predictors at the yearly and daily scales are the same, the value of their mean proportions shows significant differences. To be precise, at the yearly scale, precipitation and land cover account for up to 94.4% of the total contribution or importance to water yield, leaving the remaining 5.5% for the other four variables to share.

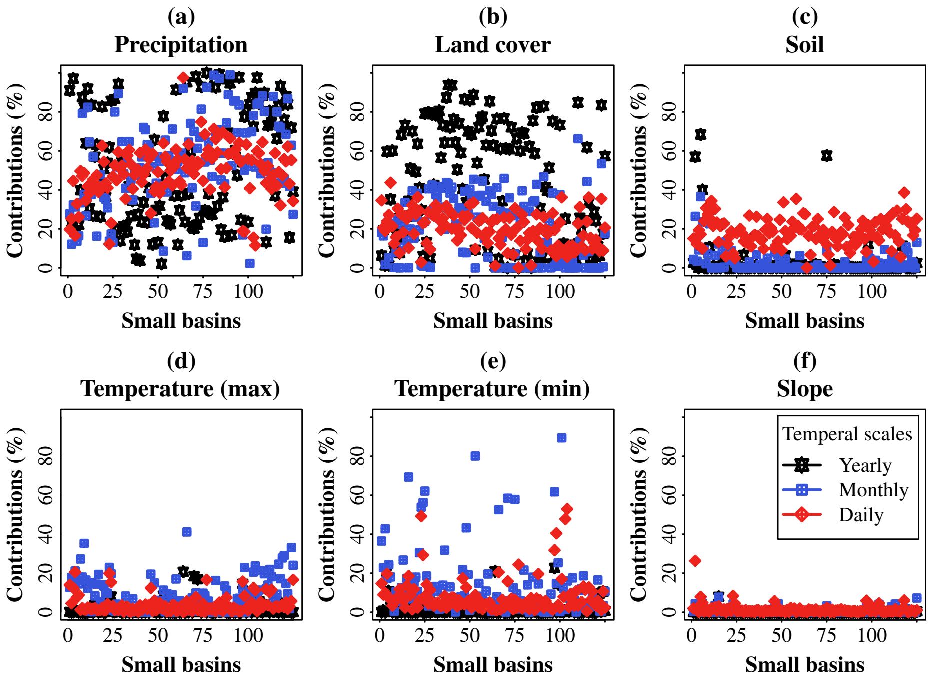

Fig. 4 Results from the BRT analysis showing the relative percent contributions of predictor variables (**a** precipitation, **b** land cover, **c** soil, **d** maximum temperature, **e** minimum temperature, and **f** slope) to water yield at the small basin scale. The number of small basins considered is 125

Table 3 Summary of BRT analysis results at the small basin scale

Variable	Yearly				Monthly				Daily				P
	μ	SD	Range	Rank	μ	SD	Range	Rank	μ	SD	Range	Rank	
Precip	52.0	29.6	2.1–100	1st	53.8	22.3	2.3–99.1	1st	49.0	13.3	11.6–97.6	1st	0.803
Land	42.5	30.4	0–93.8	2nd	19.8	16.8	0–66.3	2nd	19.8	8.8	0.2–43.7	2nd	<0.001
soil	2.5	10.1	0–68.4	3rd	2.9	5.6	0–36.6	5th	18.3	7.1	0.1–38.6	3rd	0.875
MaxTem	1.5	3.4	0–20.7	5th	9.1	8.9	0–41.1	4th	3.7	4.3	0–20.6	5th	<0.001
MinTem	1.5	4.0	0–22.7	4th	13.7	17.5	0–89.4	3rd	7.9	9.2	0.5–52.9	4th	<0.001
Slope	0.1	0.7	0–7.8	6th	0.7	1.5	0–7.7	6th	1.3	2.7	0–26.3	6th	<0.001
P	<0.001				<0.001				<0.001				

μ Mean, *SD* Standard deviation, *Precip* Precipitation, *land* Land cover, *MaxTem* Maximum temperature, *MinTem* Minimum temperature. Significances are determined using one-way ANOVA test followed by Tukey's HSD post hoc test

By contrast, at the daily scale, these four variables represent 31.2% of the total contribution. Moreover, variability, measured by the standard deviation (SD) of the contribution of each small basin (Fig. 4), tends to be large for the yearly contributions, especially for precipitation (SD = 29.6), land cover (SD = 30.4) and soil (SD = 10.1). The variability for the daily contributions is the smallest, the range for SD is between 2.7 and 13.3. Furthermore, the relative contributions at the yearly scale can be as much as 100 and 93.8 for precipitation (Fig. 4a) and land cover (Fig. 4b), respectively; while the corresponding contributions at the daily scale are 97.6 and 43.7, respectively. Note that the relative contribution of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the response. The contributions from land cover, maximum temperature, minimum temperature, and slope differ much more markedly ($P < 0.001$) among the three temporal scales compared to precipitation ($P = 0.803$) and soil ($P = 0.875$) (Table 3).

At the intermediate basin scale, the differences among the mean percent contribution for precipitation, land cover, soil and slope are not significant at any of the three temporal scales considered, as determined by the Kruskal-Wallis test (Fig. 5 and Table 4, $P > 0.05$). For instance, the percent contribution for precipitation is 65.2, 71.7 and 61.7 at the yearly, monthly, and daily scale (Table 4), respectively, and the differences among these percentages are not significant. Moreover, both the SD and range (i.e., the difference between the maximum and minimum contribution) are smaller than that of the small basin scale for all the predictors. Nevertheless, large variations in the contributions from precipitation and land cover are still evident at the intermediate basin scale.

μ Mean, *SD* Standard deviation, *Precip* Precipitation, *land* Land cover, *MaxTem* Maximum temperature, *MinTem* Minimum temperature. Significances are determined using Kruskal-Wallis tests.

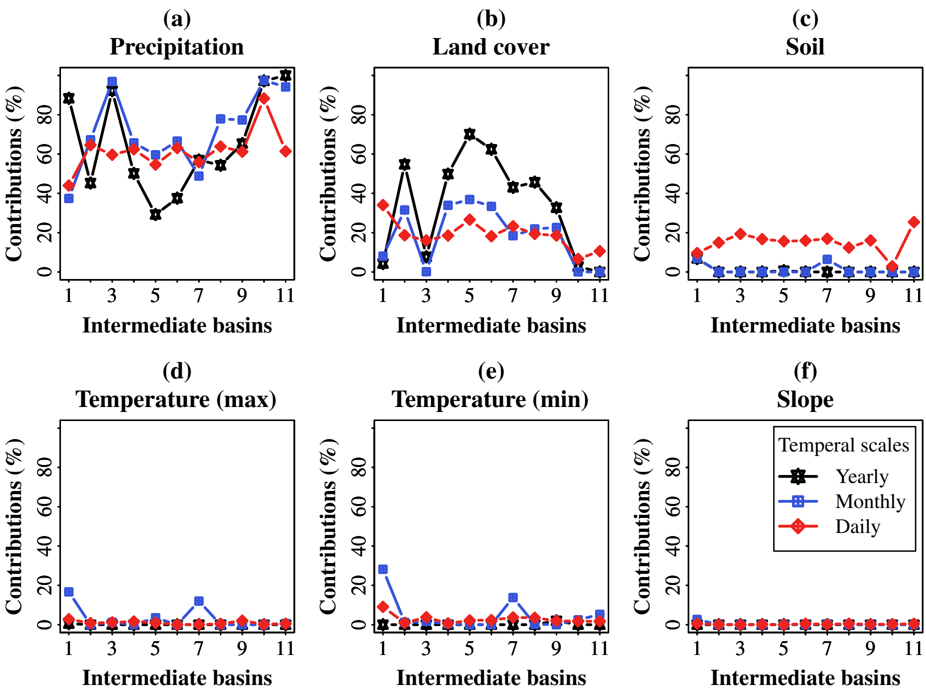

Fig. 5 Same as Fig. 4 but for the intermediate basin scale. The number of intermediate basins considered is 11

Table 4 Summary of BRT analysis results at the intermediate basin scale

Variable	Yearly				Monthly				Daily				P
	μ	SD	Range	Rank	μ	SD	Range	Rank	μ	SD	Range	Rank	
Precip	65.2	25.2	29.2–100	1st	71.7	19.5	37.4–97.5	1st	61.7	10.6	44–88.3	1st	0.633
Land	33.9	25.9	0–70.1	2nd	18.8	14.6	0–36.9	2nd	19.1	7.3	6.6–34.0	2nd	0.213
Soil	0.7	2.1	0–6.8	3rd	1.2	2.7	0–6.9	5th	15.1	5.6	2.9–25.4	3rd	0.989
MaxTem	0.0	0.2	0–0.5	5th	3.1	5.8	0–16.7	4th	1.0	0.9	0.1–2.8	5th	<0.05
MinTem	0.2	0.6	0–2.0	4th	4.8	8.8	0–28.2	3rd	2.9	2.3	0.7–9.1	4th	<0.01
Slope	0.0	0.0	0–0	6th	0.3	0.8	0–2.7	6th	0.2	0.2	0–0.5	6th	0.333
P	<0.001				<0.001				<0.001				

μ Mean, *SD* Standard deviation, *Precip* Precipitation, *land* Land cover, *MaxTem* Maximum temperature, *MinTem* Minimum temperature. Significances are determined using Kruskal-Wallis tests

Table 5 Results from the BRT analysis showing the relative percent contribution of predictors to water yield at the large basin scale. At this scale, only one large basin is considered

	Precipitation	Land cover	Soil	Min temperature	Max temperature	Slope
Yearly	68.25	31.75	0	0	0	0
Monthly	81.79	18.21	0	0	0	0
Daily	72.28	15.03	10.78	1.87	0.03	0.01

For the large basin, all of the predictor variables contribute to water yield at the daily scale (Table 5). At both the yearly and monthly scale, however, it is found that only precipitation and land cover contribute to water yield. In addition, among the three temporal scales considered, the importance of precipitation tops at the monthly scale, while the importance of land cover tops at the yearly scale.

Figure 6 summarizes the mean and SD of the contribution to water yield from the predictor variables at the different spatial and temporal scales considered. The predominance of precipitation is evident at all temporal scales followed by land cover. The contributions from the slope predictor are insignificant at all three temporal scales. At the yearly scale (Fig. 6a), the relative importance of precipitation and land cover comprise about 97.87% on average, making the influence of the other four variables negligible. A strong negative correlation is found between precipitation and temperature at the yearly scale ($R = -0.92$, $P < 0.001$, Pearson correlation). The relative importance of climate variables (e.g., precipitation,

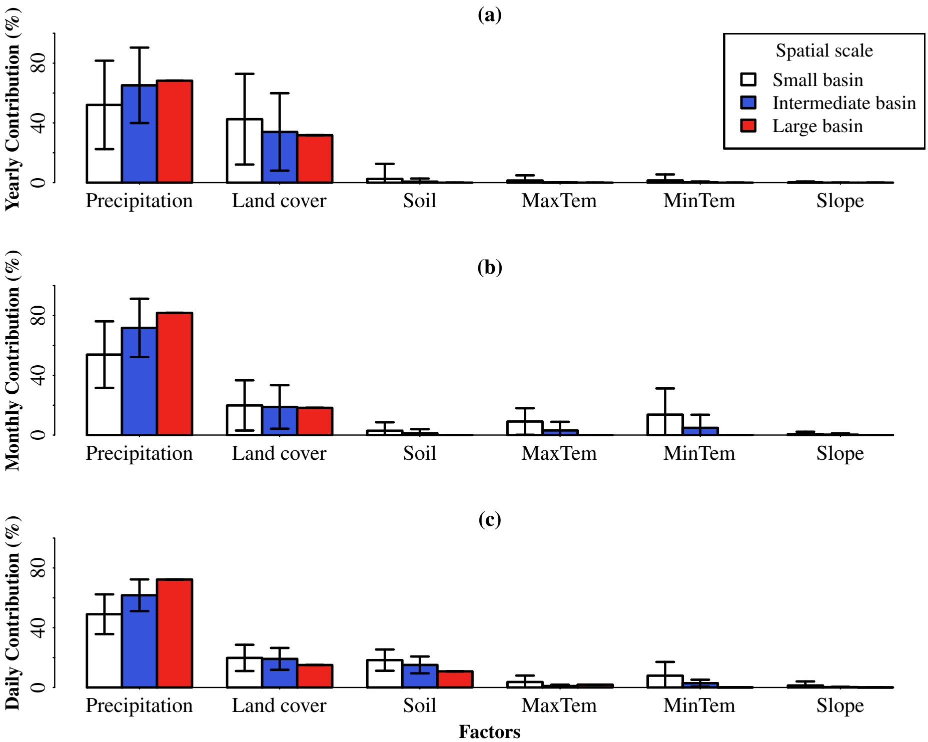

Fig. 6 Summary of the mean and SD for the BRT analysis results across spatial and temporal scales

maximum and minimum temperature) is more prominent at the monthly scale compared with both the yearly and daily scale. The largest proportion of temperature contribution occurs at the monthly scale, while soil contributions are only significant at the daily scale, which is also corroborated by Tukey's HSD post hoc tests. The yearly scale shows the most substantial variability in contribution, whereas analysis of the daily scale leads to the opposite conclusion.

The analysis presented in Fig. 6 summarizes the general behavior of the predictors according to their percent contribution to water yield. But how many small basins or intermediate basins (i.e. statistical units) follow this general behavior? Are there trends along the spatial and temporal scales? To tackle these questions, we look at each of the predictor variables and determine the total percent of basins that have a particular variable ranked within the same order (Table 6). The ranking is done at each of the temporal (yearly, monthly and daily) and spatial scales (small and intermediate basins only, since the large basin is comprised of one statistical unit). From this analysis (Table 6), it is found that the general behavior does not hold true for all the statistical units, and more statistical units tend to adhere to the proposed general behavior with increase in statistical area (i.e., spatial scale) and decrease in temporal scale. For example, at the yearly scale, precipitation ranks as the 1st most important predictor of water yield, comprising about 48.8% of the total small basins and 72.7% of the total intermediate basins. Furthermore, land cover ranks as the 2nd most important factor to contribute to water yield, consisting of about 39.2% of the total small basins whereas 63.6% of the total intermediate basins. Similar conclusions can be drawn at both the monthly and daily scales.

Table 6 Summary of the ranks of the percent contributions for the predictor variables

	Variable	Small basin						Intermediate basin					
		1st	2nd	3rd	4th	5th	6th	1st	2nd	3rd	4th	5th	6th
Year	Precip	48.8	48.8	1.6	0.8	0	0	72.7	27.3	0	0	0	0
	Land	48.0	39.2	8.0	4.8	0	0	27.3	63.6	9.1	0	0	0
	Soil	3.2	4.8	39.2	20.8	31.2	0.8	0	18.2	72.7	9.1	0	0
	MaxTem	0	3.2	48.8	36.8	10.4	0.8	0	9.1	54.5	36.4	0	0
	MinTem	0	6.4	39.2	40.8	13.6	0	0	9.1	63.6	18.2	9.1	0
	Slope	0	0.8	20	24.0	40.8	14.4	0	9.1	54.5	27.3	9.1	0
	Total	100	103.2	156.8	128.0	96.0	16.0	100	136.4	254.5	90.9	18.2	0
Month	Precip	82.4	10.4	4.8	1.6	0.8	0	100	0	0	0	0	0
	Land	4.8	48.0	10.4	22.4	10.4	4.0	0	63.6	0	36.4	0	0
	Soil	2.4	0	8.0	37.6	51.2	0.8	0	0	18.2	45.5	36.4	0
	MaxTem	0	24.8	32.8	36.8	5.6	0	0	0	63.6	36.4	0	0
	MinTem	10.4	16.8	47.2	17.6	8.0	0	0	36.4	54.5	9.1	0	0
	Slope	0	0	1.6	24.0	36.0	38.4	0	0	18.2	45.5	9.1	27.3
	Total	100	100	104.8	140	112.0	43.2	100	100	154.5	172.7	45.5	27.3
Day	Precip	91.2	3.2	4.8	0	0.8	0	100	0	0	0	0	0
	Land	2.4	46.4	38.4	7.2	4.8	0.8	0	81.8	18.2	0	0	0
	Soil	0.8	45.6	41.6	8.8	3.2	0	0	18.2	81.8	0	0	0
	MaxTem	0	1.6	1.6	16.8	63.2	16.8	0	0	0	18.2	45.5	36.4
	MinTem	4.8	3.2	13.6	64.0	12.8	1.6	0	0	0	81.8	18.2	0
	Slope	0.8	0	0	3.2	16.0	80	0	0	0	0	36.4	63.6
	Total	100	100	100	100	100.8	99.2	100	100	100	100	100	100

Precip Precipitation, *land* Land cover, *MaxTem* Maximum temperature, *MinTem* Minimum temperature. The rankings of relative contribution with the same value of 0 are all assigned to the first sequence

3.2 Marginal Effects of Predictor Variables

The BRT model for the large basin, at daily scale, is selected to detect and assess the marginal effects of individual predictors. Marginal effects are informative for summarizing how the change in water yield response is related to a change in the predictor based on a single-variable split regression tree, which conditions out the dependency of the response on the remaining predictors (De'ath 2007). Specifically, the marginal effect of each predictor on water yield is obtained here by holding all the other variables at their average values. Monotonic transformation is used in the BRT models to simplify the optimization process without affecting the marginal effect ranking orders of the input predictors. Marginal effects from the fitted models indicate that, when other variables are held constant, high water yields are most likely to be

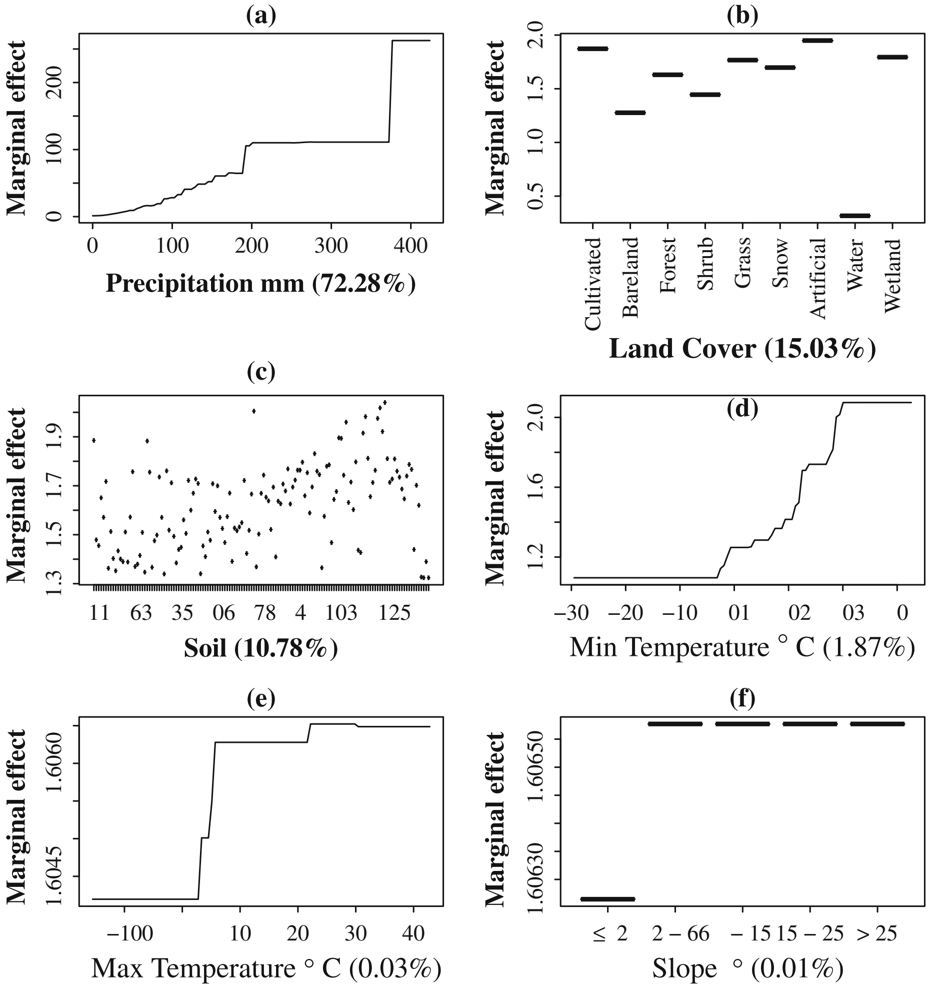

Fig. 7 Marginal effect plots for six predictor variables for water yield in the BRT model. For an explanation of variables and their units see Table 2. Note that the y-axes are on log scale

found in places with heavy precipitation, high temperature, cultivated or artificial land surface (Fig. 7).

4 Discussion

4.1 Controls on Water Yield

Not surprisingly, our analyses clearly demonstrate that precipitation, which accounts on average for one half (small basin scale) to three-fourths (large basin scale) of the relative contributions to water yield, is a driver of paramount importance for water yield across temporal and spatial scales in the YRB. It is well known that precipitation is a primary driver of water yield (Pessacg et al. 2015); its average value is often used as an indicator for water managers and policy makers to guide allocations of water resources. In the YRB, daily water yield responds positively to daily precipitation from 0 mm to around 200 mm (Fig. 7a). A stepwise respond occurs thereafter due to the fact that the only 10-year average precipitation intensity values higher than 200 mm/day are 323.7 and 423.8 mm/day.

Land cover, the second most influential predictor of water yield in this study, as indicated before by Liu et al. (2011), have average importance of 36.05, 18.94, and 17.98% at the yearly, monthly, and daily scale, respectively. At the daily scale, water yields from artificial land (featured by impervious surfaces) and cultivated land (featured by tremendous human disturbance) are the highest, followed by wetland, grassland, snow, forest, shrubland and bareland. The influence of land cover changes on water yield is a current hot topic of discussion for researchers. Water yield was found to be highly sensitive to land use land cover variation in Patagonian watershed, which typically includes some kind of gradient including forests, grasslands and shrublands (Pessacg et al. 2015). Various characteristic factors related to vegetation variations could influence water yield, including ET rates (Lu et al. 2013), water consumption (Salemi et al. 2012), water retention capacity (Wang et al. 2014), plant rooting depths (Sun et al., 2006), and percolation (Nie et al. 2011), among others.

The importance of soil on water yield tends to be more accentuated at the daily scale than at monthly and yearly scale. As indicated by Sun et al. (2006), soil thickness is an important physical basin parameter that determines how soil water used by different vegetation covers will vary due to distinct plant rooting depths. In our case, the differences in contributions of soil categories can be attributed to that different hydrologic soil groups have different soil moisture contents (Stone et al. 2003) and runoff potentials (Wang et al. 2010).

The BRT models captures well the dynamic correlation between water yield and maximum/minimum temperature at the monthly time step. Temperature is a crucial variable of the hydrologic cycle, whose importance is only second to precipitation among the climate variables (Stone et al. 2001). At yearly and daily scale, however, the contribution of temperature to water yield in the YRB is relatively low. This is consistent with other studies that the effects of precipitation could mask the effects of temperature on water yield (Lu, 2013). With this study, we find that daily water yield responds positively to daily minimum temperature from −3 °C to around 20 °C (Fig. 7d). When temperature goes beyond 20 °C (Fig. 7d), the influence of daily minimum temperature on water yield stays stable. Although the influence of daily maximum temperature on daily water yield is relatively small, according to its percentage contribution of 0.03% (Fig. 7e), a slight increase around 3 °C can be seen. Both snowfall and snow melt are associated with temperatures around 0 °C, thus both the maximum (Fig. 7e) and minimum (Fig. 7d) temperature show increasing trends after 0 °C. Interestingly, we find that minimum temperature plays a more important role than the maximum temperature in influencing water yield, albeit marginally. This could be attributed to that snowfall and snow melt exert more influence on minimum temperature than maximum temperature. More evidence should be drawn to validate this finding.

We find that the contribution of slope tends to be around zero for most cases, which indicates that compared with other environmental factors, the importance of slope in regulating water yield is in this case negligible. This may be because slope has a direct effect on the travel time of water but only an indirect impact on total water yield. Further research is needed to provide additional evidence for this.

4.2 Implications and Recommendations for Future Studies

Researchers have found that the relationship between basin physical properties and water yield is variable (Lu et al. 2013; Wang et al. 2011). However, research efforts dedicated to uncovering changes in water yield response across both spatial and temporal scales are rare. In this study, by combining the SWAT with BRT, the relationships between water yield and predictor variables at multiple spatial and temporal scales are quantitated and compared. Our analyses clearly demonstrate that the response of water yield to climate and basin characteristics is geographically variable and largely dependent on the scale.

In terms of spatial scales, distinct characteristics are observed from the large basin to small basin scales. On one hand, the macroscale basin perspective gives a whole picture of the general relationship between water yield and its determinants along with their marginal effects, which is especially beneficial for macroscopic risk prevention and high-level decision-making. On the other hand, significant differences in relative contributions are found within smaller scale basins (Fig. 3 and Table 6). Clearly, the impact of climate and basin characteristics on water yield is basin-specific, making it impossible to specify the actual quantitative relationships using a universally applicable criterion. In this context, we suggest that it is inappropriate to discuss the contribution of different environmental factors to water yield without referring to the specific site and temporal scale. Moreover, the large variations in water yield response greatly complicate extrapolation across scales, suggest that any extrapolation of water yield relationships from one basin to another must be done with caution (Berg et al. 2016; de Vente and Poesen 2005; Zhou et al. 2015). The significant differences in relative contributions can be attributed to the non-linear nature of processes within the hydrological cycle.

In terms of the temporal scales considered, the results tend to become more stable and reliable from the yearly to the daily scale. Firstly, judging from the SD values, variations decrease from the yearly to daily scale. Secondly, the coarser scale may not tell the complete story because the contribution of land cover at yearly scale is very dominant making the remaining four factors insignificant, e.g., soil moisture is one of the many factors that can be assumed negligible at the yearly scale. Similarly, at the monthly scale, the importance of meteorological predictors on water yield are amplified due to their explicit seasonal character, while this is not the case at the daily scale. Finally, yet importantly, simulations at the daily scale are more detail and rich in their representation of hydrological processes than coarser scales. The daily scale offers an opportunity for studying and observing the relative contribution of factors that may be misrepresented at coarser scales. It is reasonable to recommend that researchers studying changes in water yield should prioritize conducting their studies at finer temporal scales. Previous studies have mostly been focused on the yearly scale (Brown et al. 2005), particularly for large basins.

Although BRT is advantageous in capturing the complex and non-linear relationships between hydrological respondents and their explanatory variables. The potential drawbacks of BRT models should be noted. First, the confidence interval as well as the significance of difference cannot be derived for BRT's relative contribution calculation (Chung 2013). Second, the BRTs require extensive calculations to optimize parameters of improved predictive performance, including shrinkage, tree complexity, number of trees, and bagging fraction, which escalates the computation time (Saha et al. 2015).

5 Conclusions

This study analyzes the dependence and contribution of predictors to water yield in the YRB, with the aim of eventually informing and supporting more effective water resources management goals. A new method coupling a hydrological model, SWAT, with a statistical-machine learning model, BRT, is implemented to study the dependence of water yield on multiple environmental variables. A few key environmental variables including precipitation, maximum and minimum temperature, land cover, soil, and slope are selected because they have previously been shown to contribute to the total water yield. Different basin (i.e., small, intermediate and large) and temporal (i.e., daily, monthly, yearly) scales are employed for understanding how water yield varies with scale. Based on the results obtained, the main finding of this study can be summarized as follows:

- Among the six predictors considered, precipitation and land cover are the most important factors while slope has the lowest importance. The relative importance of the remaining predictors differs among temporal scales. Soil is the third most important contributor to water yield at both the yearly and daily scales, while minimum and maximum temperature surpass soil at the monthly scale due to their explicit seasonal character. The average contributions obtained from the BRT models could serve as useful indicators for water managers and policy makers in guiding water resource allocations for the YRB.
- In terms of the marginal effects of predictors, daily water yield responds positively to daily precipitation from around 0 to 200 mm. Among different land cover types, artificial land and cultivated land contribute the most to water yield. Daily water yield responds positively to daily minimum temperature from −3 °C (around the snow melting point) to around 20 °C, while maximum temperature shows gradient characteristics.
- The relationship between environmental factors and water yield is site- and scale-specific. For the small basin scale, precipitation and land cover combined account for 94.4% and 68.8% of the total contribution to water yield at the yearly and daily scale, respectively. By contrast, for the intermediate basin scale, the differences among the mean percent contribution for precipitation, land cover, soil and slope are not significant at any of the three temporal scales considered. For the large basin scale, all of the predictor variables contribute to water yield at the daily scale, however, only precipitation and land cover contribute to water yield at both the yearly and monthly scale. Therefore, we suggest that caution should be taken when extrapolating water yield relationships across spatial and temporal scales.

– In contrast with results derived from small basins, the relationship between water yield and its determinants obtained from the macroscale basin perspective may be particularly beneficial for macroscopic risk prevention and high-level decision-making. More stable and reliable results can be derived at the daily scale compared to the yearly. Previous studies have mostly been focused on the yearly scale, we recommend that researchers studying changes in water yield should prioritize conducting their studies at multiple spatial and temporal scales, and include finer temporal scales.

Acknowledgments This study is supported by the Major Science and Technology Program for Water Pollution Control and Treatment (Grant No. 2017ZX07207003-01).

References

Adams KN, Fowler AM (2006) Improving empirical relationships for predicting the effect of vegetation change on annual water yield. J Hydrol 321:90–115

Arnold JG, Kiniry JR, Srinivasan R, Williams JR, Neitsch SL (2012) Soil and Water Assessment Tool Theoretical Documentation. https://swat.tamu.edu/media/99192/swat2009-theory.pdf. Accessed June 2019

Berg MD, Marcantonio F, Allison MA, McAlister J, Wilcox BP, Fox WE (2016) Contrasting watershed-scale trends in runoff and sediment yield complicate rangeland water resources planning. Hydrol Earth Syst Sci 20:2295–2307

Brown AE, Zhang L, McMahon TA, Western AW, Vertessy RA (2005) A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. J Hydrol 310:28–61

Castillo CR, Güneralp İ, Güneralp B (2014) Influence of changes in developed land and precipitation on hydrology of a coastal Texas watershed. Appl Geogr 47:154–167

Chung YS (2013) Factor complexity of crash occurrence: an empirical demonstration using boosted regression trees. Accid Anal Prev 61:107–118

De'ath G (2007) Boosted trees for ecological modeling and prediction. Ecology 88:243–251

de Vente J, Poesen J (2005) Predicting soil erosion and sediment yield at the basin scale: Scale issues and semi-quantitative models. Earth Sci Rev 71:95–125

Elith J, Leathwick J (2015) Boosted Regression Trees for ecological modeling. https://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf. Accessed June 2019

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802–813

Farley KA, Jobbagy EG, Jackson RB (2005) Effects of afforestation on water yield: a global synthesis with implications for policy. Glob Chang Biol 11:1565–1576

Feng XM, Sun G, Fu BJ, Su CH, Liu Y, Lamparski H (2012) Regional effects of vegetation restoration on water yield across the Loess Plateau, China. Hydrol Earth Syst Sci 16:2617–2628

Friedman JH, Meulman JJ (2003) Multiple additive regression trees with application in epidemiology. Stat Med 22:1365–1381

Gharun M, Vervoort RW, Turnbull TL, Adams MA (2014) A test of how coupling of vegetation to the atmosphere and climate spatial variation affects water yield modelling in mountainous catchments. J Hydrol 514:202–213

Gokbulak F, Sengonul K, Serengil Y, Ozhan S, Yurtseven I, Uygur B, Ozcelik MS (2016) Effect of forest thinning on water yield in a sub-humid mediterranean oak-beech mixed forested watershed. Water Resour Manag 30:5039–5049

Golden HE, Lane CR, Prues AG, D'Amico E (2016) Boosted regression tree models to explain watershed nutrient concentrations and biological condition. J Am Water Resour Assoc 52:1251–1274

Hale R, Marshall S, Jeppe K, Pettigrove V (2014) Separating the effects of water physicochemistry and sediment contamination on Chironomus tepperi (Skuse) survival, growth and development: A boosted regression tree approach. Aquat Toxicol 152:66–73

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. https://link.springer.com/content/pdf/10.1007/978-1-4614-7138-7.pdf. Accessed June 2019

Kuria FW, Vogel RM (2015) Global storage-reliability-yield relationships for water supply reservoirs. Water Resour Manag 29:1591–1605

Liu J, Sui C, Deng D, Wang J, Feng B, Liu W, Wu C (2016) Representing conditional preference by boosted regression trees for recommendation. Inf Sci 327:1–20
Liu Y, Leng X, Deng Z, Wang L, Zhang L, Liu S, An S (2011) Effects of watershed vegetation on tributary water yields during the wet season in the Heishui Valley, China. Water Resour Manag 25:1449–1464
Lu N, Sun G, Feng XM, Fu BJ (2013) Water yield responses to climate change and variability across the North-South Transect of Eastern China (NSTEC). J Hydrol 481:96–105
McDonnell J et al (2007) Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. Water Resour Res 43
Mehta VM, Rosenberg NJ, Mendoza K (2011) Simulated Impacts of Three Decadal Climate Variability Phenomena on Water Yields in the Missouri River Basin1. J Am Water Resour Assoc 47:126–135
Naghibi SA, Pourghasemi HR, Abbaspour K (2018) A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. Theor Appl Climatol 131:967–984
Neitsch SL, Kiniry JG, Williams JR (2009) Soil And Water Assessment Tool: theoretical documentation. https://swat.tamu.edu/media/99192/swat2009-theory.pdf. Accessed June 2019
Nie W, Yuan Y, Kepner W, Nash MS, Jackson M, Erickson C (2011) Assessing impacts of Landuse and Landcover changes on hydrology for the upper San Pedro watershed. J Hydrol 407:105–114
Pessacg N, Flaherty S, Brandizi L, Solman S, Pascual M (2015) Getting water right: A case study in water yield modelling based on precipitation data. Sci Total Environ 537:225–234
Price K (2011) Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. Prog Phys Geogr 35:465–492
R Core Team (2014) The R project for statistical computing. https://www.r-project.org. Accessed June 2019
Rice JS, Emanuel RE, Vose JM, Nelson SAC (2015) Continental U.S. streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics. Water Resour Res 51:6262–6275
Ridgeway G (2007) Generalized Boosted Models: A guide to the gbm package. http://www.saedsayad.com/docs/gbm2.pdf. Accessed June 2019
Saha D, Alluri P, Gan A (2015) Prioritizing Highway Safety Manual's crash prediction variables using boosted regression trees. Accid Anal Prev 79:133–144
Salazar F, Toledo MA, Onate E, Suarez B (2016) Interpretation of dam deformation and leakage with boosted regression trees. Eng Struct 119:230–251
Salemi LF, Groppo JD, Trevisan R, de Moraes JM, Lima WD, Martinelli LA (2012) Riparian vegetation and water yield: A synthesis. J Hydrol 454:195–202
Stone MC, Hotchkiss RH, Hubbard CM, Fontaine TA, Mearns LO, Arnold JG (2001) Impacts of climate change on Missouri Rwer Basin water yield. J Am Water Resour Assoc 37:1119–1129
Stone MC, Hotchkiss RH, Mearns LO (2003) Water yield responses to high and low spatial resolution climate change scenarios in the Missouri River Basin. Geophys Res Lett 30
Sun F, Kuang W, Xiang W, Che Y (2016) Mapping water vulnerability of the Yangtze River Basin: 1994–2013. Environ Manag 58:857–872
Sun G, Zhou G, Zhang Z, Wei X, McNulty SG, Vose JM (2006) Potential water yield reduction due to forestation across China. J Hydrol 328:548–558
van Dijk A, Pena-Arancibia JL, Bruijnzeel LA (2012) Land cover and water yield: inference problems when comparing catchments with mixed land cover. Hydrol Earth Syst Sci 16:3461–3473
Wang G, Yang H, Wang L, Xu Z, Xue B (2014) Using the SWAT model to assess impacts of land use changes on runoff generation in headwaters. Hydrol Process 28:1032–1042
Wang S, Fu B, He C, Sun G, Gao G (2011) A comparative analysis of forest cover and catchment water yield relationships in northern China. For Ecol Manag 262:1189–1198
Wang X, Shang S, Yang W, Clary CR, Yang D (2010) Simulation of land use–soil interactive effects on water and sediment yields at watershed scale. Ecol Eng 36:328–344
Wu F, Zhan J, Chen J, He C, Zhang Q (2015) Water yield variation due to forestry change in the head-water area of Heihe River basin, Northwest China. Adv Meteorol 2015:1–8
Xu J, Yang D, Yi Y, Lei Z, Chen J, Yang W (2008) Spatial and temporal variation of runoff in the Yangtze River basin during the past 40 years. Quat Int 186:32–42
Yao Y, Cai T, Ju C, He C (2015) Effect of reforestation on annual water yield in a large watershed in northeast China. J For Res 26:697–702
Zhang Q, Xu C, Zhang Z, Chen Y, Liu C, Lin H (2008) Spatial and temporal variability of precipitation maxima during 1960–2005 in the Yangtze River basin and possible association with large-scale circulation. J Hydrol 353:215–227
Zhou G et al (2015) Global pattern for the effect of climate and land cover on water yield. Nat Commun 6:5918

Affiliations

Fengyun Sun[1,2,3] **· Alfonso Mejia**[4] **· Yue Che**[1,2,3]

[1] School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China

[2] Shanghai Key Lab for Urban Ecological Processes and Eco-Restoration, East China Normal University, Shanghai 200241, China

[3] Institute of Eco-Chongming (IEC), Shanghai 200062, China

[4] Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA 16802, USA