AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2017WR021144

Key Points:
- Spatial streamflow correlation is predicted using a novel stochastic analytical model based on catchment-scale hydrological properties
- How different hydrological processes affect streamflow similarities is assessed
- A method to estimate streamflow correlation in absence of discharge data is developed that is used to identify hydrologically similar sites

Correspondence to:
A. Betterle,
andrea.betterle@eawag.ch

Citation:
Betterle, A., Radny, D., Schirmer, M., & Botter, G. (2017). What do they have in common? Drivers of streamflow spatial correlation and prediction of flow regimes in ungauged locations. *Water Resources Research*, *53*, 10,354–10,373. https://doi.org/10.1002/2017WR021144

Received 19 MAY 2017
Accepted 15 NOV 2017
Accepted article online 27 NOV 2017
Published online 10 DEC 2017

What Do They Have in Common? Drivers of Streamflow Spatial Correlation and Prediction of Flow Regimes in Ungauged Locations

A. Betterle[1,2,3], D. Radny[1], M. Schirmer[1,2], and G. Botter[3]

[1]Department of Water Resources and Drinking Water, EAWAG (Swiss Federal Institute of Aquatic Science and Technology), Dübendorf, Switzerland, [2]The Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Neuchâtel, Switzerland, [3]Department ICEA and International Center for Hydrology "Dino Tonini", University of Padova, Padua, Italy

Abstract The spatial correlation of daily streamflows represents a statistical index encapsulating the similarity between hydrographs at two arbitrary catchment outlets. In this work, a process-based analytical framework is utilized to investigate the hydrological drivers of streamflow spatial correlation through an extensive application to 78 pairs of stream gauges belonging to 13 unregulated catchments in the eastern United States. The analysis provides insight on how the observed heterogeneity of the physical processes that control flow dynamics ultimately affect streamflow correlation and spatial patterns of flow regimes. Despite the variability of recession properties across the study catchments, the impact of heterogeneous drainage rates on the streamflow spatial correlation is overwhelmed by the spatial variability of frequency and intensity of effective rainfall events. Overall, model performances are satisfactory, with root mean square errors between modeled and observed streamflow spatial correlation below 10% in most cases. We also propose a method for estimating streamflow correlation in the absence of discharge data, which proves useful to predict streamflow regimes in ungauged areas. The method consists in setting a minimum threshold on the modeled flow correlation to individuate hydrologically similar sites. Catchment outlets that are most correlated ($\rho > 0.9$) are found to be characterized by analogous streamflow distributions across a broad range of flow regimes.

1. Introduction

The spatial and temporal variability of streamflows critically impact the life of riverine biomes and the security of anthropogenic water uses (Postel & Richter, 2003; Sabo et al., 2010; Widder et al., 2014). Therefore, understanding the physical processes that shape hydrological regimes across and along river basins represents a scientific problem with relevant socio-economic implications, including the development of strategies aimed at restoring riverine habitats and at the design of optimal configurations of water infrastructures.

The spatial correlation of daily streamflows represents an effective and synthetic index that quantitatively encapsulates the similarity between the hydrographs at two arbitrary outlets (Archfield & Vogel, 2010). Analogous discharge dynamics result from the spatial correlation of climatic and hydrological properties in the relevant contributing areas (Skøjen et al., 2006). Characterizing hydrologic similarity between catchments enables a deeper understanding on how heterogeneity of the underlying geomorphic and climatic drivers is propagated through the hydrologic cycle and eventually affects spatial patterns of flow regimes (e.g., Doulatyari et al., 2017; Schaefli et al., 2014; Lahaa et al., 2014). Exploring the spatial structure of streamflows has been suggested to play an important role in a number of fields, ranging from the expansion of existing hydrometric networks (Chacon-Hurtado et al., 2017; Messinger & Paybins, 2014) to the identification of spatial patterns of ecological variables along stream networks (Isaak et al., 2014; Mc Guire et al., 2014). Furthermore, it has been shown that streamflow correlation represents a better surrogate than spatial proximity to individuate river sections characterized by analogous flow dynamics (Archfield & Vogel, 2010). As a consequence, in regionalization techniques aimed at predicting streamflow regimes at ungauged outlets (see, e.g., Blöschl et al., 2013), streamflow correlation could represent both a useful index to individuate target (ungauged) and donor (gauged) sites, and a metric to classify or rank catchments based on similarities of flow dynamics.

Recently, Betterle et al. (2017) have developed a stochastic, physically based approach to characterize the spatial correlation of daily discharges. The approach exploits a parsimonious framework with reduced complexity that focuses on how the stochastic nature of rainfall propagates through the catchment-scale water balance to flow regimes (Botter et al., 2007a, 2007b). Analytical expressions were identified that link the streamflow correlation to the frequency and intensity of effective rainfall events in the contributing catchments, and to the recession properties of the hydrograph. River network topology is properly accounted for in that framework, since the effect on streamflow correlation resulting from the nested (or nonnested) nature of catchments is considered in the definition of model parameters (Müller & Thompson, 2015). Thanks to the underlying hypothesis on streamflow formation and routing, the method causally relates flow dynamics at two river sites with simple climatic and landscape features of the corresponding drainage basins. Existing statistical and geostatistical methods used to estimate spatial patterns of flow statistics, instead, typically overlook the physical description of runoff generation processes. Therefore, though versatile and practice oriented, these methods prevent a direct link between rainfall and streamflow dynamics and they are often challenged by the quantity and quality of the available data (Blöschl et al., 2013; Müller & Thompson, 2015; Skøjen & Blöschl, 2007). The analytical structure of the model proposed by Betterle et al. (2017), combined with its parsimonious nature, allows a simple assessment of the sensitivity of flow dynamics to different hydrological drivers, making the formulation applicable to a wide range of settings with reduced computational efforts and low data requirements. Therefore, its use in sparsely gauged or ungauged regions is an appealing avenue of research, still largely unexplored.

In this paper, the framework proposed by Betterle et al. (2017) is utilized to investigate how observed intercatchment heterogeneity in the underlying physical processes (e.g., rainfall, runoff production, drainage rates) impact the spatial variability of flow regimes. The method allows one to disentangle the effect of intertwined climatic and landscape features on the spatial patterns of flow dynamics. The main research hypothesis is that spatial gradients of climatic properties bear a fundamental and recognizable signature on the cross-correlation of daily flows, which can be efficiently used to predict streamflow regimes in the absence of discharge data. This research hypothesis is tested through a detailed application of the stochastic approach to a set of catchments in a $75,000$ km^2 region of the Eastern United States, where synchronous daily rainfall and discharge data are available.

The specific goals of the study can be summarized as follows: (i) testing the performance of the analytical approach, combined to different parameter estimation procedures, using observed streamflow and rainfall data; (ii) quantifying the main heterogeneity of climatic and hydrologic attributes among the study catchments and evaluating the impact of such heterogeneity on streamflow correlation; (iii) developing and testing a method for predicting streamflow correlations in settings where discharge time series are not available; and (iv) provide a proof of concept that the method can be used to predict streamflow regimes in the absence of discharge measurements.

The paper is organized as follows: section 2 summarizes the analytical framework used to perform the analysis; section 3 presents the case studies and the hydrologic data. Section 4 discusses two alternative approaches to estimate the parameters of the model. Section 5 discusses the performances of the models and relevant results. Section 6 concludes the paper.

2. Methods

The steady state spatial correlation of daily streamflows at the outlet of two arbitrary catchments is studied using a parsimonious physically based approach. The method is applicable to catchments not affected by relevant water storages (lakes, reservoirs, or snowpacks) and where streamflow dynamics can be causally linked to precipitation. The model is suited to catchments where the typical response time of the stream network is shorter than 1 day (say, catchments with sizes up to about $10,000$ km^2) (Betterle et al., 2017). The linear, zero lag cross-correlation of daily discharge is analytically derived from the moment generating function of the joint discharge dynamics at two selected outlets. The geomorphic and climatic features of the contributing areas are captured by a parsimonious set of catchment-scale parameters that express the frequency and intensity of effective rainfall events (i.e., the streamflow generating rainfall events), as well as the recession rates in the two catchments. The theoretical framework relies on a number of simplifying assumptions, which are summarized in the following.

Daily streamflow dynamics within each catchment are seen as sequences of abrupt increases of discharge (streamflow jumps) followed by recessions (Botter et al., 2007a; Claps et al., 2005; Doulatyari et al., 2015). Streamflow recessions in between jumps are assumed to be exponential with rate k (i.e., $1/k$ is the average catchment response time) (Botter et al., 2007c, 2013; Müller & Thompson, 2015; Pumo et al., 2013). Streamflow jumps are seen as the response to effective rainfall events taking place in the contributing catchment, where effective rainfall is defined as the fraction of rainfall that exceeds the water holding capacity of the soil. Soil water dynamics during wetting-drying cycles are controlled by the water storage capacity of the soil (i.e., soil porosity, rooting depth, field capacity) as well as vegetation and climatic factors (e.g., evapotranspiration, precipitation). The effective rainfall is therefore dependent on precipitation features and the antecedent moisture of the catchment, which in turn is a function of soil, vegetation, and climate (Botter et al., 2013; Milly, 1994; Porporato et al., 2004; Thompson et al., 2011). The effective rainfall within each contributing area is modeled according to a Poisson process of frequency λ_{it}, where the subscript i identifies one of the two relevant catchments ($i \in \{1, 2\}$) and t denotes the "total" sequence of effective rainfall events (all effective rainfall events in the catchment). The corresponding effective rainfall depths are assumed to be exponentially distributed random variables with mean α_{it}.

The streamflow correlation between two selected outlets is then studied by decoupling the "total" sequence of effective rainfall events into two independent subsets (see Figure 2 in Betterle et al. (2017)): (i) effective rainfall events that occur simultaneously in the two catchments (joint events) and (ii) effective rainfall events that occur in only one of the two catchments (disjoint events). Joint and disjoint events are modeled as independent Poisson processes and described in terms of mean frequency and mean effective rainfall depth. In particular, λ_{12} and λ_i represent, respectively, the frequency of joint effective rainfall events and the frequency of disjoint events in the catchment i. Likewise, the distribution of the depths of joint and disjoint events is assumed to be different, and the average effective rainfall depths in the catchment i during joint and disjoint events are denoted as α_i^{12} and α_i, respectively. For a detailed summary of all model parameters, the reader is directed to Table 1.

In this paper, three alternative analytical expressions for the spatial correlation between the streamflow time series at two catchment outlets (ρ_{model}) are considered (Betterle et al., 2017):

$$\rho_{model}^{(1)} = \overbrace{\overbrace{\frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}}}^{F_{\lambda}^{(1)}} \overbrace{\frac{1}{2}(1+r_{\alpha})}^{F_{\alpha}^{(1)}}}^{F_{\lambda\alpha}^{(1)}} \overbrace{\frac{2\sqrt{k_1 k_2}}{k_1+k_2}}^{F_{k}^{(1)}} \tag{1}$$

$$\rho_{model}^{(2)} = \overbrace{\frac{\lambda_{12}\alpha_1^{12}\alpha_2^{12}}{\sqrt{\left[\lambda_1(\alpha_1)^2+\lambda_{12}(\alpha_1^{12})^2\right]\left[\lambda_2(\alpha_2)^2+\lambda_{12}(\alpha_2^{12})^2\right]}} \frac{1}{2}(1+r_{\alpha})}^{F_{\lambda\alpha}^{(2)}} \overbrace{\frac{2\sqrt{k_1 k_2}}{k_1+k_2}}^{F_{k}^{(2)}} \tag{2}$$

$$\rho_{model}^{(3)} = \overbrace{\overbrace{\frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}}}^{F_{\lambda}^{(3)}} \overbrace{\frac{\alpha_1^{12}\alpha_2^{12}}{\alpha_{1t}\alpha_{2t}} \frac{1}{2}(1+r_{\alpha})}^{F_{\alpha}^{(3)}}}^{F_{\lambda\alpha}^{(3)}} \overbrace{\frac{2\sqrt{k_1 k_2}}{k_1+k_2}}^{F_{k}^{(3)}} \tag{3}$$

where r_{α} represents the correlation between joint effective rainfall depths in the considered catchments. These expressions correspond to different assumptions on the depth distribution of effective rainfall events. Equation (1) assumes that the effective rainfall depths of joint and disjoint events are described by the same exponential probability density function (i.e., all effective rainfall events are described by the same depth distribution). Equation (2) assumes that joint and disjoint effective rainfall depths are characterized by two different exponential probability density functions (PDF) with means α_i^{12} and α_i, respectively. Equation (3) assumes two independent exponential PDFs for joint and total effective rainfall depths, with means α_i^{12} and α_{it}, respectively. Therefore, equations (2) and (3) allow one to account for the effect of the diverse physical processes that are possibly involved in the generation of joint and disjoint events (e.g., large-scale

Table 1
Summary of the Parameters

Parameter	Description
λ_{it}	Average frequency of all effective rainfall events in catchment *i*
λ_i	Average frequency of disjoint effective rainfall events in catchment *i*
λ_{12}	Average frequency of joint effective rainfall events in the two catchments
λ_m	Minimum between λ_{1t} and λ_{2t}
α_{it}	Average depth of all effective rainfall events in catchment *i*
α_i	Average depth of disjoint effective rainfall events in catchment *i*
α_i^{12}	Average depth of joint effective rainfall events in catchment *i*
$\tilde{\lambda}_{it}$	Average frequency of all rainfall events in catchment *i*
$\tilde{\lambda}_i$	Average frequency of disjoint rainfall events in catchment *i*
$\tilde{\lambda}_{12}$	Average frequency of joint rainfall events in the two catchments
$\tilde{\alpha}_{it}$	Average depth of all rainfall events in catchment *i*
$\tilde{\alpha}_i$	Average depth of disjoint rainfall events in catchment *i*
$\tilde{\alpha}_i^{12}$	Average depth of joint rainfall events in catchment *i*
r_α	Correlation between the joint effective rainfall depths in the two catchments
$r_{\tilde{\alpha}}$	Correlation between the joint rainfall depths in the two catchments
k_i	Streamflow decay rate during recessions in catchment *i*

Note. Joint: refers to events occurring on the same day in both catchments; disjoint: refers to events occurring on different days in the two catchments; $i = 1,2$ identifies the two catchments.

versus local convective or orographic events). Equation (1) represents the simplest model, and it can be interpreted as a special case of equation (2) and (3) for $\alpha_i = \alpha_i^{12}$ and $\alpha_{it} = \alpha_i^{12}$, respectively.

Equations (1)–(3) can ideally be decomposed into a number of factors that identify the contribution to streamflow correlation associated to different hydrological processes, namely: (i) the relative frequency of joint effective rainfall events (F_λ); (ii) the intensity of effective rainfall events (F_α), and (iii) the features of recession rates (F_k). Note that in equation (2) the effect of the frequency and intensity of events is blended into a single term, $F_{\lambda\alpha}^{(2)}$. The superscripts of the terms F_λ, F_α, F_k, and $F_{\lambda\alpha}$ refer to the specific solution considered (equations (1)–(3)).

Note that the model is able to account for the topological arrangement of the considered catchments, as the nested or nonnested nature of the two catchments is reflected in the frequency of joint effective rainfall events (λ_{12}). Indeed, in case of nested catchments, $\lambda_{12} = \min\{\lambda_{1t}, \lambda_{2t}\} = \lambda_{1t}$ (the subscript 1 indicates here the inner catchment). The runoff produced from any nested subcatchment of a river indeed propagates along the network and affects the streamflow dynamics in downstream sites. In case of nonnested catchments, on the other hand, the frequency of joint events is typically smaller than the minimum runoff frequency in the two catchments: $\lambda_{12} \leq \min\{\lambda_{1t}, \lambda_{2t}\}$. More details on the derivation of the model equations and hypothesis are given in Betterle et al. (2017).

3. Case Studies and Hydrologic Data

The analytical model (equations (1)–(3)) is applied to 13 nested and nonnested catchments in a $75,000\ km^2$ region of the eastern US (Table 2). The size of the study catchments, which are situated in North Carolina, Virginia, and Tennessee, spans from about 100 to 5,000 km^2 (Figure 1). All basins are weakly impacted by natural or artificial water storages (reservoirs, lakes), and they are characterized by diversified geomorphoclimatic and landscape features. All the study catchments belong to the MOPEX data set (http://www.nws.noaa.gov/ohd/mopex/, Schaake et al., 2006), which provides an adequate record of hydrological variables including rainfall and streamflows. In particular, streamflow time series at the outlet of each catchment, as well as mean areal daily rainfall records from 1948 to 2003 were used in this study. Streamflow data included in the MOPEX data set are provided by USGS (https://waterdata.usgs.gov/nwis/rt), whereas spatially averaged daily rainfall rates are computed based on the PRISM model (http://www.prism.oregonstate.edu) (Daly et al., 2008). Minimum and maximum average values of mean annual precipitation, mean annual streamflow, altitude, and slope across the study catchments are, respectively: $p_{min-max} = 3.0-5.4\ mm/d$, $q_{min-max} = 1.3-3.8\ mm/d$, $h_{min-max} = 480-1220\ m$, $s_{min-max} = 1.8$–6.4%. Land cover in the study area is rather heterogeneous and varies from evergreen needle leaf forest (0–50%) to deciduous

Table 2
Summary Information About the Study Catchments

Number	USGS code	Name	Area (km^2)	Min–Max altitude (m a.s.l.)	Streamflow gauging station	State
1	03532000	Powell River	1,774	328–1,263	Arthur	TN
2	03531500	Powell River	826	394–1,263	Jonesville	VA
3	03528000	Clinch River	3,818	335–1,424	Tazewell	TN
4	03490000	Holston River	1,738	370–1,442	Gate City	VA
5	03504000	Nantahala River	134	958–1,643	Rainbow Springs	NC
6	03550000	Valley River	269	482–1,532	Tomotla	NC
7	03455000	French Broad River	4,812	317–1,925	Newport	TN
8	03451500	French Broad River	2,448	570–1,925	Asheville	NC
9	03448000	French Broad River	1,751	608–1,808	Bent Creek	NC
10	03443000	French Broad River	767	634–1,808	Blantyre	NC
11	03465500	Nolichucky River	2,085	469–2,002	Embreeville	TN
12	03470000	Little Pigeon River	914	273–1,998	Sevierville	TN
13	03512000	Oconaluftee River	477	573–1,877	Birdtown	NC

broadleaf forests (20–90%), grassland and cropland (0–15%) (University of Maryland land use classification; Hansen et al., 2000). Despite the pronounced interseasonal variability of the mean discharge in the study area, mainly induced by monthly patterns of potential ET along the year, there are no prolonged dry periods and significant carryover effects between seasons in the region. Location and extent of the study area were selected based on the appropriate density of stations with long-term synchronous streamflow and rainfall records, and because of the pronounced geomorphoclimatic gradients observed in the region, which allow a broad range of seasonal flow regimes and streamflow correlations to be explored.

From the 13 case studies, 78 combinations of catchment pairs can be identified. The analysis is carried out at seasonal timescale, with seasons defined based on fixed calendar dates (spring: March, April, May; summer: June, July, August; autumn: September, October, November; and winter: December, January, February). This leads to 312 couples of seasonal streamflow correlations (given N catchments, the number of possible pairs of outlets is $\frac{N(N-1)}{2}$). For each season, the streamflow (Pearson) correlation coefficient between the discharge time series observed at two generic outlets 1 and 2 (ρ_{meas}) can be calculated as:

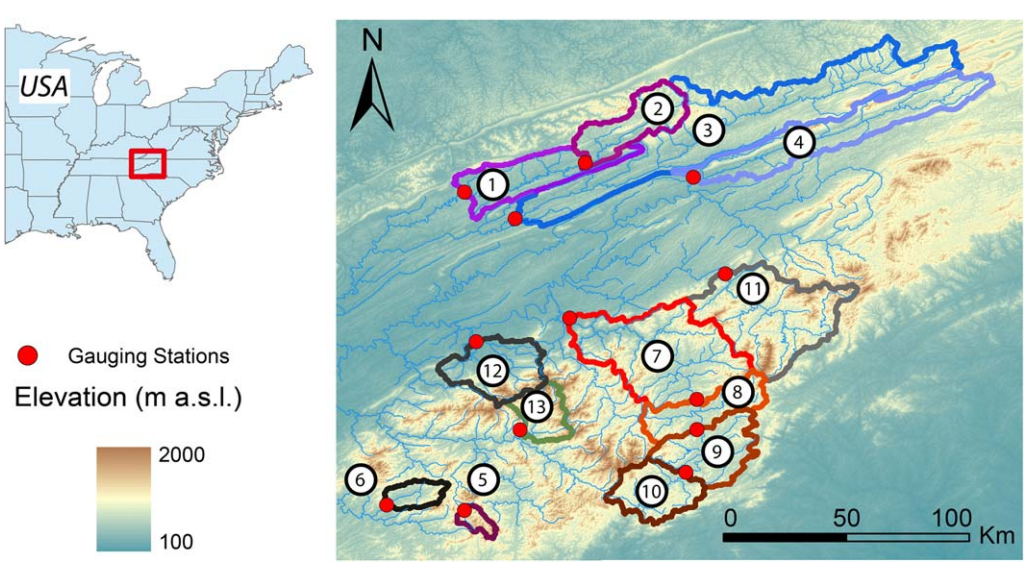

Figure 1. The study area includes 13 nested and nonnested catchments located in the eastern United States.

$$\rho_{meas} = \frac{\sum_{j=1}^{n} [(q_1(j) - \langle q_1 \rangle)(q_2(j) - \langle q_2 \rangle)]}{\sqrt{\sum_{j=1}^{n} (q_1(j) - \langle q_1 \rangle)^2 \sum_{j=1}^{n} (q_2(j) - \langle q_2 \rangle)^2}} \quad (4)$$

where $q_1(j)$ and $q_2(j)$ are the streamflow at the outlets 1 and 2 at day j, n is the number of recorded days, and $\langle q_1 \rangle$ and $\langle q_2 \rangle$ are the sample averages of q_1 and q_2. Equation (4) is then used to evaluate model performances as discussed in what follows.

4. Parameter Estimation

In the application of the proposed framework, a key issue is represented by the estimation of the model parameters. Two procedures, which make use of different types of hydrologic data, are described in the following. Alternative procedures could be potentially identified depending on the actual data availability, a feature which makes the proposed framework flexible, and applicable to different contexts.

4.1. Method A: Estimate of Model Parameters From Streamflow Time Series

According to the model formulation, each effective rainfall event produces a discontinuity in the hydrograph (i.e., an abrupt increase of discharge). The frequency of effective rainfall events (λ_{it}, λ_{i}, and λ_{12}) can therefore be inferred from the observed number of jumps in the daily streamflow records at the relevant outlets. The observed jumps are first classified as "joint" or "disjoint" according to their timing (joint events correspond to synchronous daily jumps in both catchments). Then, the frequency of total, joint, and disjoint events is calculated by dividing the number of recorded events of each type by the duration of the considered time series. Similarly, the average effective rainfall intensity, α, can be evaluated from the magnitude of the daily streamflow jumps. Since an exponential unit hydrograph is assumed by the model, the depth of each effective rainfall pulse $h(j)$ can be computed from the correspondent flow increment $\Delta Q(j)$ as $h(j) = \frac{\Delta Q(j)}{k}$ (see, e.g., equation (1) in Betterle et al. (2017) and equation (4) in Botter et al. (2007a)). Consequently, $\alpha = \frac{\langle \Delta Q(j) \rangle}{k}$. The analysis is carried out separately for each type of streamflow-producing events (namely total, joint, and disjoint), enabling the estimate of the corresponding mean depths (α_{it}, α_{j}^{12}, and α_{i}, respectively). In addition, r_α is estimated as the Pearson correlation coefficient between the joint effective rainfall depths in the two catchments. Finally, the recession rates k_1 and k_2 are evaluated from the observed hydrographs by means of recession analysis. Since the model assumes exponential recessions, the drainage rate k_i is estimated by fitting a linear regression on different pairs $(\frac{\Delta Q}{\Delta t}, Q)$ selected from the descending limbs of observed hydrographs in the corresponding catchment (Basso et al., 2015; Ceola et al., 2010; Dralle et al., 2015).

One of the main advantages of estimating the model parameters based on observed discharge data is that the effect of soil moisture dynamics is implicitly accounted for. Indeed, streamflow dynamics portray the signature of catchment-scale soil moisture variability, which in turn encapsulate the effect of climatic and landscape attributes (e.g., Laio et al., 2001; Milly, 1994; Rodriguez-Iturbe et al., 1999; Settin et al., 2007). Nevertheless, this method cannot be applied to catchments where streamflow data are lacking. Moreover, estimating frequency and intensity of effective rainfall events from streamflow time series may be cumbersome, and the result can be sensitive to the specific algorithm adopted to identify streamflow jumps and recessions from discharge records (Cheng & Krajewski, 2016).

4.2. Method B: Estimate of Model Parameters in the Absence of Discharge Data

This estimation method is rooted in the idea that frequency and intensity of effective rainfall events are strongly dependent on rainfall dynamics. As noted earlier, the frequency of effective rain is smaller than the precipitation frequency because some rainfall inputs are buffered by soil moisture dynamics in the root zone. However, particularly in cases where soil and vegetation features do not show pronounced heterogeneity, the intercatchment variability of rainfall properties is likely to be a primary control on the spatial variability of the frequency and intensity of effective rainfall events. Additionally, the main driver of the streamflow correlation is represented by rainfall events leading to synchronous streamflow jumps at multiple outlets (joint events), which may be triggered by large-scale and intense humid fronts, as observed in the study area (Messinger & Paybins, 2014). Therefore, heterogeneity in the buffer capacity of the soil is

likely to bear a reduced effect on the spatial variability of the frequency and magnitude of intense effective rainfall events, limiting the impact of soil and vegetation heterogeneity on the flow correlation.

Based on these arguments, we assume that the terms F_λ, F_α, and $F_{\lambda\alpha}$ in equations (1), (2), and (3) can be calculated based on the rainfall data, instead of effective rainfall time series. More specifically, in equation (1) we assume that:

$$\frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}} \approx \frac{\tilde{\lambda}_{12}}{\sqrt{\tilde{\lambda}_{1t}\tilde{\lambda}_{2t}}} \tag{5}$$

and in equation (2)

$$\frac{\lambda_{12}\alpha_1^{12}\alpha_2^{12}}{\sqrt{\left[\lambda_1(\alpha_1)^2+\lambda_{12}(\alpha_1^{12})^2\right]\left[\lambda_2(\alpha_2)^2+\lambda_{12}(\alpha_2^{12})^2\right]}} \approx \frac{\tilde{\lambda}_{12}\tilde{\alpha}_1^{12}\tilde{\alpha}_2^{12}}{\sqrt{\left[\tilde{\lambda}_1(\tilde{\alpha}_1)^2+\tilde{\lambda}_{12}(\tilde{\alpha}_1^{12})^2\right]\left[\tilde{\lambda}_2(\tilde{\alpha}_2)^2+\tilde{\lambda}_{12}(\tilde{\alpha}_2^{12})^2\right]}} \tag{6}$$

where the tilde ($\sim$) denotes the average frequencies and intensities of rainfall events in the two catchments. In particular, the parameters $\tilde{\lambda}_{it}$ and $\tilde{\alpha}_{it}$ are the mean frequency and the mean depth of the total spatially averaged daily rainfall time series in the ith contributing catchment. Likewise, $\tilde{\lambda}_{12}$ and $\tilde{\alpha}_i^{12}$ are the mean frequency and depth of joint rainfall events (i.e., rain events that bring nonzero precipitation in the two catchments on the same day), while $\tilde{\lambda}_i$ and $\tilde{\alpha}_i$ are the analogous statistics for disjoint rainfall events. In addition, the correlation r_α between the intensity of joint effective rainfall time series in equations (1)–(3) is estimated as the correlation between the corresponding joint rainfall events $r_{\tilde{\alpha}}$. In the same vein, in equation (3), we also assume that $\frac{\alpha_1^{12}\alpha_2^{12}}{\alpha_{1t}\alpha_{2t}} \approx \frac{\tilde{\alpha}_1^{12}\tilde{\alpha}_2^{12}}{\tilde{\alpha}_{1t}\tilde{\alpha}_{2t}}$, which means that the ratio between the product of the average joint effective rainfall depths and the product of the average total effective rainfall depths is assumed to be the same as the corresponding ratio calculated based on the rainfall depths. The latter assumption is supported by theoretical arguments, according to which the distribution of effective rainfall depths is weakly impacted by soil moisture dynamics in case of exponentially distributed rain depths (Verma et al., 2011). The same assumption has allowed robust estimates of flow duration curves under a wide range of climatic, landscape, and vegetation features (Basso et al., 2015; Botter et al., 2007c, 2013; Doulatyari et al., 2015; Mejia et al., 2014; Müller et al., 2014; Müller & Thompson, 2015; Pumo et al., 2013).

Finally, since the estimate of recession properties is challenging in the absence of discharge data (Biswal & Marani, 2014; Doulatyari et al., 2015), and because of the limited impact of heterogeneity in recession properties on the spatial correlation of streamflows (Betterle et al., 2017), we assume for simplicity that $\frac{2\sqrt{k_1 k_2}}{k_1+k_2}=1$ for all the selected pairs of outlets. More refined methods for the estimate of F_k can be developed depending on the need of each study (e.g., Dralle et al., 2015).

The parameter estimation procedure described in this section does not account for soil moisture dynamics. However, when model parameters are directly estimated from rainfall records, the likely overestimation of frequency and intensity of both joint and disjoint events can bring a limited influence on the ratios involved in the definition of the terms F_λ, F_α, and $F_{\lambda\alpha}$ (equations (1)–(3)), therefore reducing the sensitivity of the modeled correlation to possible biases induced by the parameter estimation procedure. Additionally, the evaluation of some model parameters (e.g., λ_{12}) is facilitated by the use of synchronous rainfall records, which allows an easier assessment of the timing of rainy days and the corresponding rainfall depths.

Given the importance of an accurate estimate of the spatiotemporal gradients of rainfall, reliable information on the spatial patterns of daily rainfall are a prerequisite for the successful application of the method. This requires the availability of rainfall records at multiple rainfall gauges, properly interpolated by means of geostatistical techniques (e.g., kriging) or physiographical methods (e.g., PRISM). Note that, a 1 mm threshold on the spatially averaged daily rainfall over the relevant catchments has been applied in order to account for canopy interception (Doulatyari et al., 2017; Lai & Katul, 2000; Laio et al., 2001).

5. Results and Discussion

5.1. Prediction of Streamflow Spatial Correlation and Its Seasonality

The observed correlation coefficient of daily streamflows at different pairs of outlets belonging to the selected case studies (equation (4)) is compared to the estimates provided by the analytical

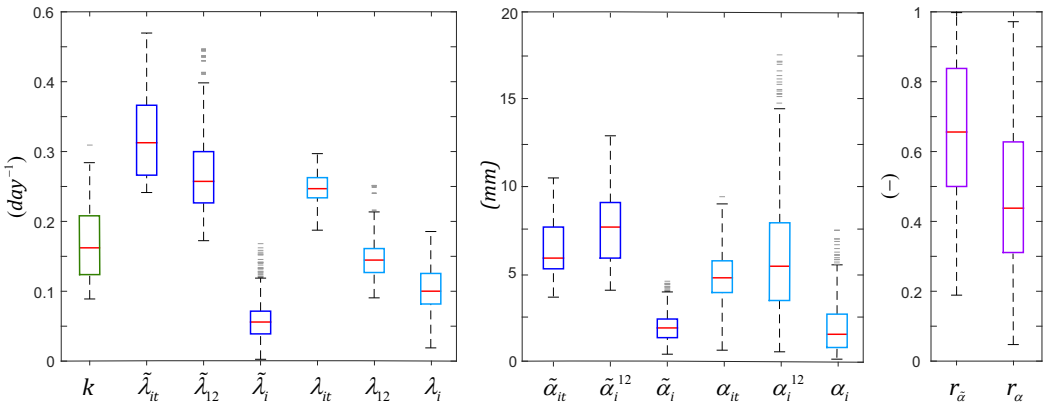

Figure 2. The boxplots on the left show the distribution of the streamflow recession rates (k), the frequency of total, joint, and disjoint rainfall events ($\tilde{\lambda}_t$, $\tilde{\lambda}_{12}$, $\tilde{\lambda}_i$) and the frequency of total, joint, and disjoint effective rainfall events (λ_t, λ_{12}, λ_i). In the middle are the total, joint, and disjoint effective rainfall depths estimated from rainfall ($\tilde{\alpha}_{it}, \tilde{\alpha}_i^{12}, \tilde{\alpha}_{it}$) and from streamflow records ($\alpha_{it}, \alpha_i^{12}, \alpha_i$). On the right is shown the correlation of the joint effective rainfall depths (r_α) and the correlation of the joint rainfall depths ($r_{\tilde{\alpha}}$).

model—equations (1–3)—using different parameter estimation procedures (sections 4.1 and 4.2). The distribution of the model parameters across the case studies is summarized in the boxplots in Figure 2. The variability of the recession parameter k suggests heterogeneous drainage characteristics in the study area, which includes catchments featured by different morphological characteristics in terms of size, shape, relief, and land cover. The distribution of rainfall and effective rainfall frequencies is also shown in Figure 2. As expected, the overall frequency of events (λ_{it}, $\tilde{\lambda}_{it}$) is larger than the frequency of joint and disjoint events. Effective rainfall frequencies are smaller than the corresponding precipitation frequencies because of the soil water deficit created by evapotranspiration. Additionally the figure shows that effective rainfall intensities estimated from rainfall records ($\tilde{\alpha}_{it}, \tilde{\alpha}_i^{12}, \tilde{\alpha}_i$) are comparable to those estimated directly from streamflow records ($\alpha_{it}, \alpha_i^{12}, \alpha_i$). Figure 2 also displays the distribution of the correlation coefficient between joint effective rainfall depths (r_α) and joint rainfall depths ($r_{\tilde{\alpha}}$). Joint rainfall depths are more correlated than joint effective rainfall depths ($r_{\tilde{\alpha}} > r_\alpha$), and $r_{\tilde{\alpha}}$ exhibits a reduced intercatchment variability. This is likely to be the effect of the heterogeneity of soil and vegetation features across the study catchments.

Model performance is graphically assessed by means of scatterplots where the observed and modeled streamflow correlations for the 312 pairs of seasonal discharge time series are compared. A quantitative estimation of the model performance is also provided via the root mean square error (*RMSE*) between modeled and observed correlations.

Figure 3 represents the model performance when parameters are estimated from discharge time series as discussed in section 4.1. Equations (2) (Figure 3b) and equation (3) (Figure 3c) capture reasonably well the observed variability of streamflow correlation across the study catchments. Conversely, equation (1) (Figure 3a) systematically underestimates the observed correlations because it disregards the difference between the intensity of joint and disjoint effective rainfall events. The underestimation is enhanced for intermediate values of ρ_{meas}, corresponding to catchments pairs where the frequencies of joint and disjoint events are comparable. In this range, disregarding differences in the intensities of joint and disjoint events and assigning potentially overestimated effective depths to disjoint events, can cause a visible decrease of $\rho_{model}^{(1)}$. On the contrary, $\rho_{model}^{(1)}$ is less biased for pairs of catchments characterized by high (or low) correlations, because they experience mainly one of the two classes of events (i.e., joint or disjoints).

Among the three alternatives models, equation (2) is the best performer (*RMSE* = 0.094), followed by equation (3). Note that joint events can be difficult to individuate and quantify from streamflow time series, leading to underestimated values λ_{12} or biased estimates of α_i^{12}. This could explain the slight underestimation of $\rho_{model}^{(2)}$, that can however be addressed by including more sophisticated algorithms for streamflow time series analysis. It is also worth noting that the performance of equation (2) further increases when the

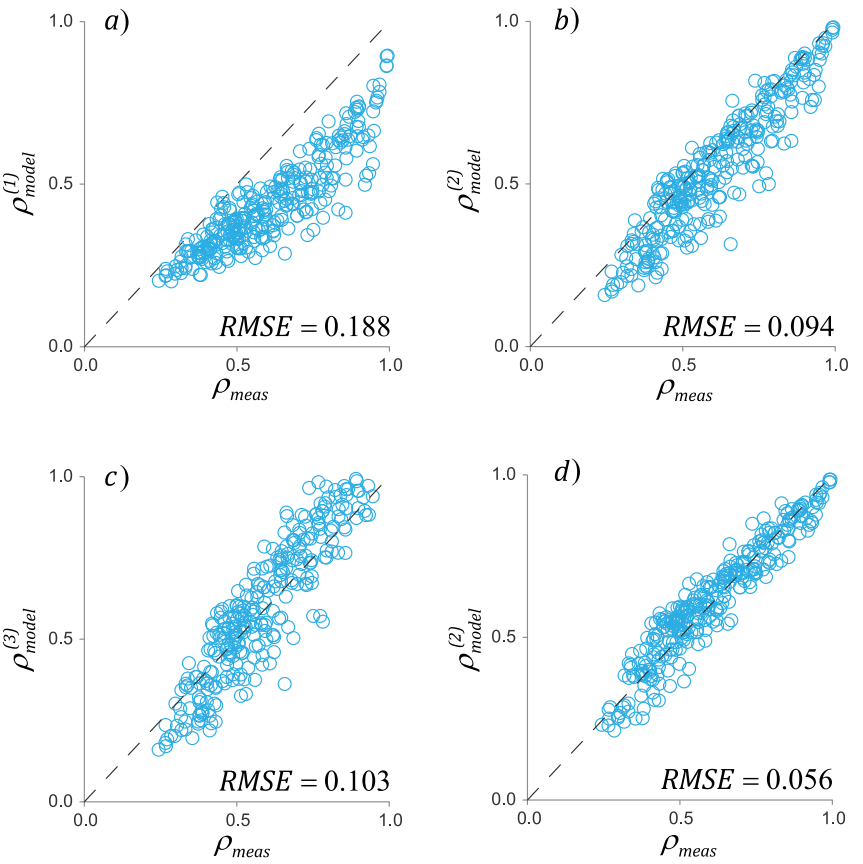

Figure 3. The scatterplots compare the seasonal streamflow correlation calculated from daily discharge records (*x* axis) and the streamflow correlation estimated by the analytical model (equations (1)–(3)) when the parameters are obtained from streamflow records (*y* axis) for the 312 pairs of seasonal streamflow time series between the 13 study catchments. Scatterplots (a), (b), and (c) refer to equations (1), (2), and (3) respectively. Scatterplot (d) shows the performances of equation (2) when the intensities of joint effective rainfall events are estimated directly from rainfall records.

intensity of effective events is directly estimated from rainfall time series (Figure 3d), with a *RMSE* equal to 0.056. Overall, the superior performances of equation (2) can be explained by the larger flexibility of this version of the model, where the different intensity of joint and disjoint events is properly accounted for.

Figure 4 shows the results obtained by means of equations (1), (2), and (3) when the parameters are estimated without exploiting discharge data, as discussed in section 4.2. As expected, a general decrease of performance is observed. In particular, the higher correlation of joint rainfall events—if compared to the corresponding correlation of effective rainfall events, see Figure 2—leads to overestimated streamflow correlations in all versions of the model. In this case, the simplest version of the model, represented by equation (1) (Figure 4a), benefits more than the others from the use of rainfall-estimated model parameters and outperforms equations (2) and (3). This can be explained by a more robust evaluation of synchronicity and intensity of the events combined to a simpler model structure, which results in more reliable estimations when less informative input data are available (rainfall versus streamflow). Figure 4d shows the performances of equation (1) when the assumption $F_k = 1$ is relaxed, and F_k is calculated from observed streamflow records as in Figure 3. The increase of model performance is quite small ($\approx$ 10%). This result hints at the limited impact of recession heterogeneity on the observed daily streamflow correlation in the study region.

The satisfactory performance provided by equation (1) with rainfall-estimated parameters seems to be an appealing prospect for the prediction of streamflow correlation in catchments where hydrometric stations are lacking (see section 5.3). In particular, the successful application of equation (1) highlights how the timing and intensity of joint and disjoint rainfall events is a primary factor controlling spatial patterns of streamflow correlation and should be considered in the selection of donors and receiving sites when estimating

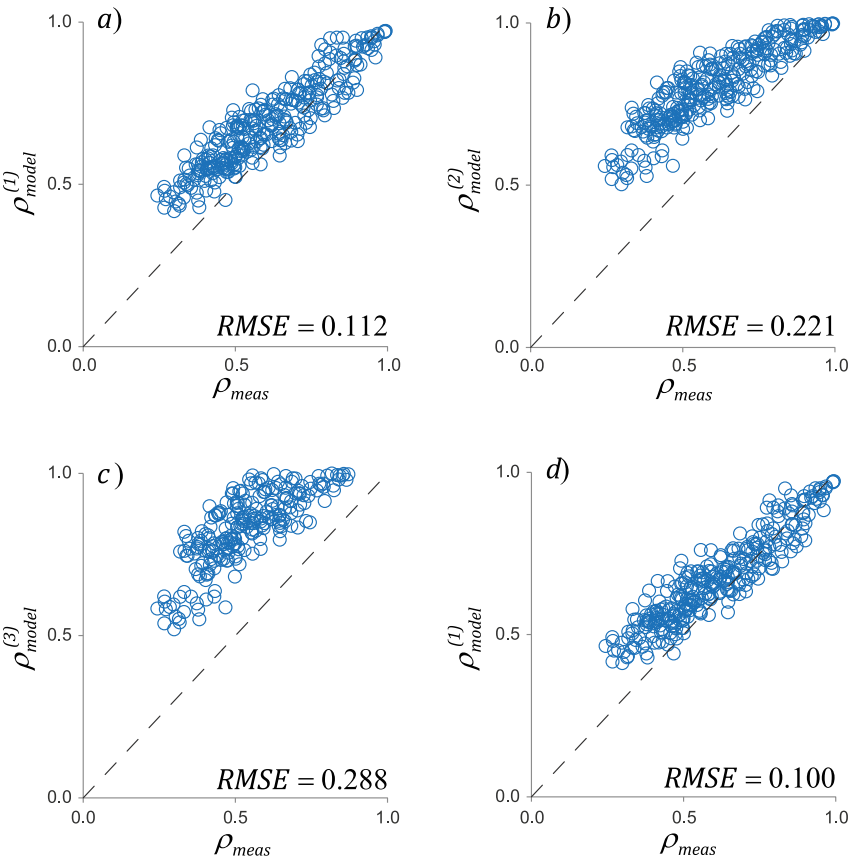

Figure 4. The scatterplots compare the streamflow correlation calculated from daily discharge records (*x* axis) and the streamflow correlation estimated by the analytical model (equations (1)–(3)) when the parameters are obtained directly from rainfall (*y* axis) for the 312 pairs of seasonal streamflow time series between the 13 study catchments. Scatterplots (a), (b), and (c) refer to equations (1), (2), and (3), respectively, assuming $F_k = 1$. Scatterplot (d) shows the performances of equation (1) when the streamflow recession parameters are estimated based on the discharge data.

streamflow characteristics at ungauged sites. The improved model performances when soil moisture dynamics are indirectly accounted for (parameters estimated based on the streamflow records as in section 4.1 and Figure 3) or when the assumption of homogeneous recession rates is relaxed (Figure 4d), suggest the existence of cases where streamflow correlation cannot be explained solely by rainfall dynamics. Nevertheless, homogeneous vegetation and morphological characteristics, possibly emerging from the coevolution of landscapes experiencing analogous climate and rainfall regimes (Huang & Niemann, 2006; Jefferson et al., 2010; Sivapalan et al., 2011), can justify the application of the framework in the absence of discharge data. Especially when the identification of highly correlated sites is concerned.

Model performances are further analyzed in Figure 5, which displays the results obtained at annual timescale using equations (1), (2), and (3), when model parameters are estimated using different methods (see caption of Figure 5 for details). The plots show good performance of the model in all cases with $RMSE < 0.10$. These results suggest that the steady state formulation of the model is able to properly incorporate the effect of the seasonal variability of the spatial heterogeneity of key hydrological processes.

The plots in Figure 6 show the patterns of streamflow correlation across the study area as a function of the distance between the center of mass of the contributing catchments. The observed trends are properly portrayed by the analytical formulation (in this case equation (2) is used with parameters estimated from streamflow data). Figure 6 shows that the distance is a strong control on streamflow correlation. However, a significant scatter and strong seasonal variability are observed, which are related to the anisotropic heterogeneity of climatic and landscape properties in the study area. Scattering and seasonal variability of the

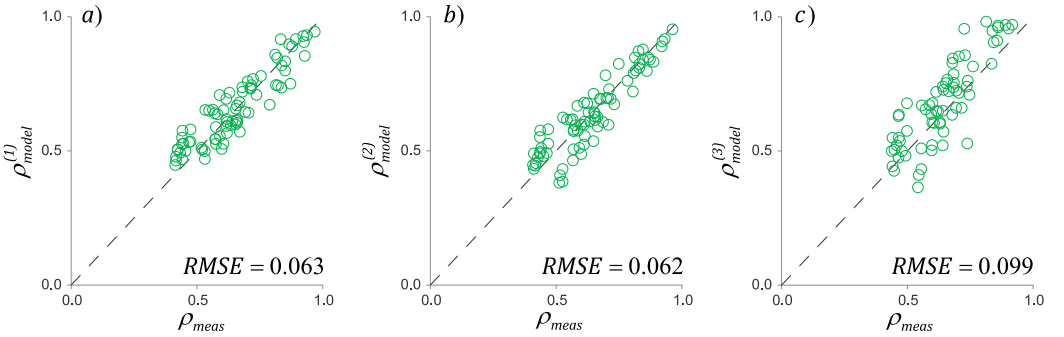

Figure 5. The scatterplots compare the streamflow correlation calculated from daily discharge records (x axis) and the streamflow correlation estimated by the analytical model (y axis) for the 78 pairs of annual streamflow time series selected from the 13 study catchments. In plot (a), streamflow correlation is estimated by means of equation (1) when F_λ and F_α are obtained from rainfall data and F_k is estimated from discharge time series. In plot (b), streamflow correlation is estimated by means of equation (2) when F_λ and F_k are obtained from streamflow records and F_α is obtained from rainfall records. In plot (c), streamflow correlation is estimated by means of equation (3) when model parameters are obtained from streamflow records.

relationship between the spatial correlation of discharges and distance seem to be adequately captured by the analytical model.

The seasonal streamflow correlation is above the annual average during spring and winter. On the contrary, summer and autumn display a sharper decrease in correlation with increasing distance. Especially during summer, intercatchment differences in terms of land cover (vegetation) and water retention capacity is possibly enhanced by higher evapotranspiration rates, increasing the intercatchment variability of the capacity to buffer incoming rainfall through soil water deficit. Moreover, spatially heterogeneous soil moisture dynamics and the presence of convective storms affecting only a limited number of sites in the study region, eventually result in less correlated streamflow time series during summer (Messinger & Paybins, 2014). During spring and winter, instead, reduced spatial heterogeneity of rainfall is responsible for a milder decrease of streamflow correlation with increasing distance. The reduced spatial heterogeneity in the hydrological response during winter can be attributed not only to the increased homogeneity of rainfall but also to enhanced runoff coefficients during this season (Doulatyari et al., 2015). On the other hand, high correlations observed in spring can be related to the presence of large humid fronts causing frequent and intense joint effective rainfall events over the entire study area.

The shift of seasonal streamflow correlation along the year in relation to model performances is further investigated in Figure 7, where equation (2) is employed. A general shift toward more correlated streamflows is observed in winter and spring, whereas a general loss of correlation, and a visible increase in the intercatchment variability is observed in autumn and summer. Performances are satisfactory across all the four season (see also boxplots in Figure 7), with a slight tendency of the model to underestimate streamflow correlation during summer. Instead, model performance tend to increase during spring ($RMSE < 0.5$).

Variations of streamflow spatial correlation along the year are a consequence of the intertwined seasonal variability of climatic forcings and vegetation dynamics. Seasonal changes of average precipitation and evapotranspiration, together with changes in their spatial autocorrelation structure, can lead to concurrent effects on streamflow dynamics and possibly affect connectivity patterns along river networks during the year. From a practical perspective, the quantification of the seasonal variability of the streamflow spatial correlation can provide hints for an optimal use of water resources (e.g., hydropower, irrigation) through comprehensive management plans developed at regional scales.

5.2. Analysis of the Impact of Intercatchment Variability of Hydroclimatic Features on the Streamflow Spatial Correlation

The analytical model allows an assessment of how different climatic and hydrologic factors influence streamflow correlation. The histograms in Figure 8 represent the frequency distribution of the factors $F_{\lambda\alpha}$ and F_k in equations (1) and (2) for the 312 pairs of case studies. Given the multiplicative nature of these equations, values of F approaching 1 indicate a limited impact on the correlation of daily flows. Figure

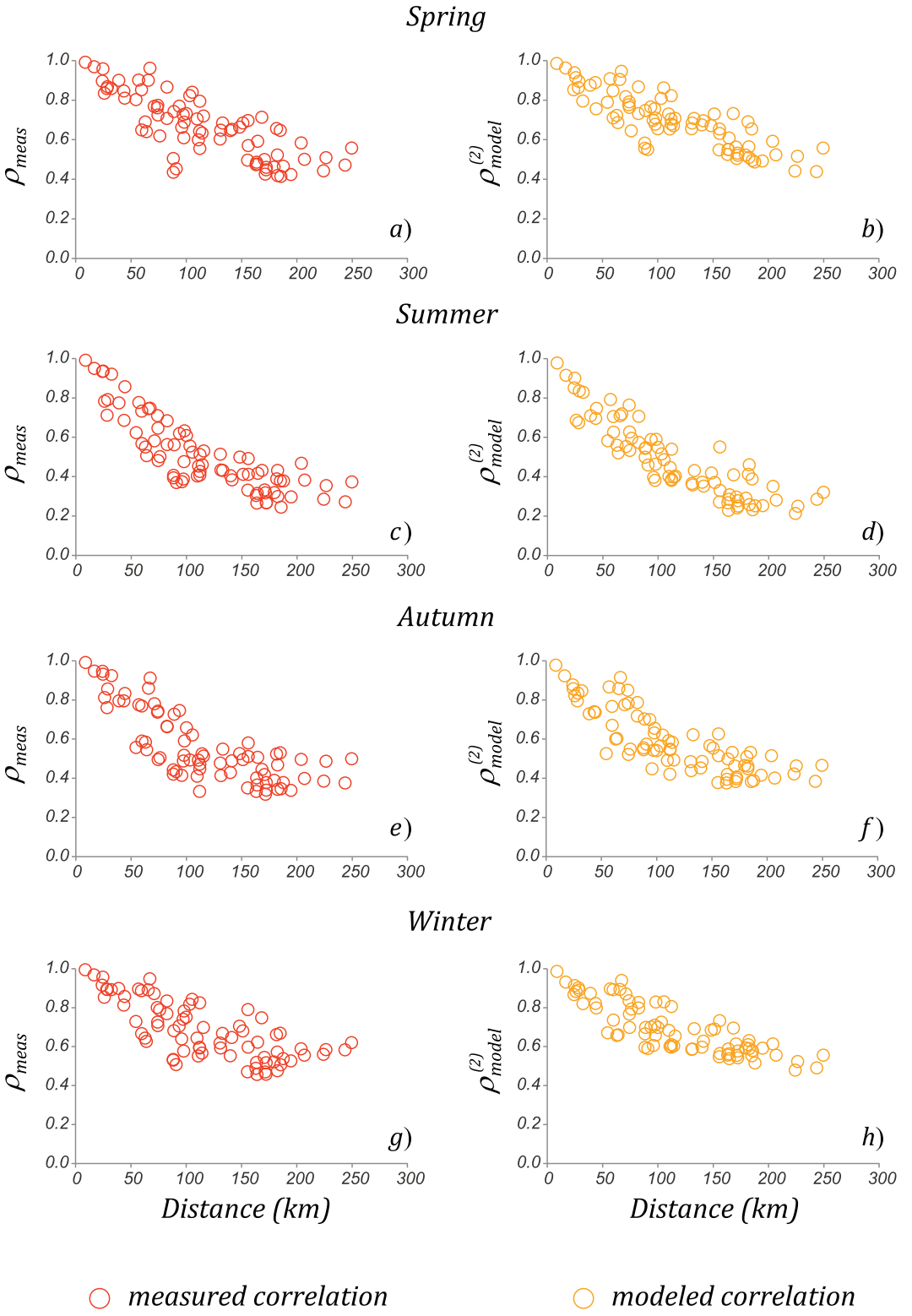

Figure 6. The seasonal streamflow correlation decreases with increasing catchment distance (distance measured from the centroids of the relevant basins). The figure shows the observed streamflow correlation and the analytical streamflow correlation obtained using equation (2) with parameters estimated from discharge as a function of the intercatchment distance.

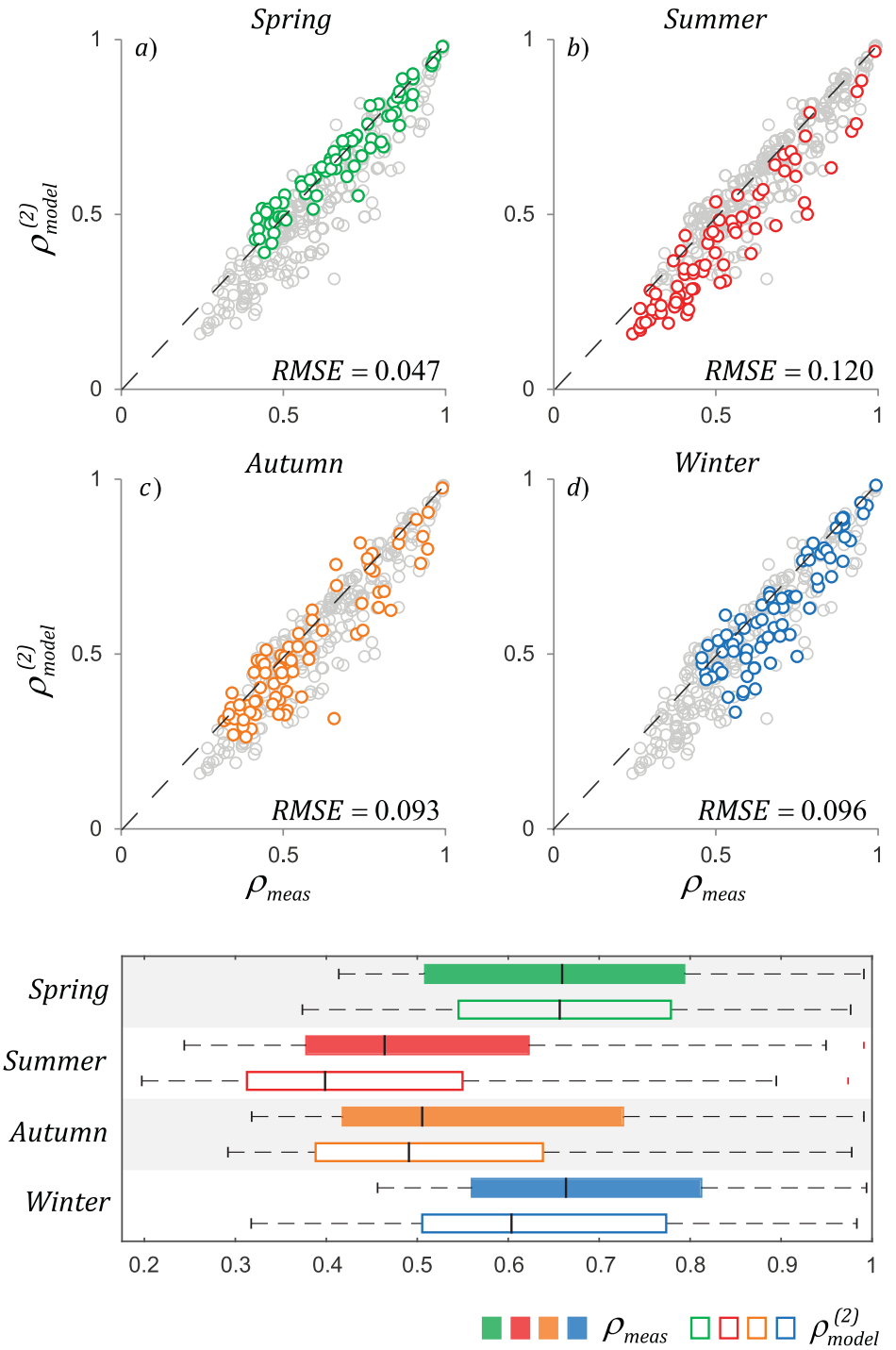

Figure 7. The scatterplots show the seasonal shift of streamflow correlation throughout the year. Summer and autumn display generally lower and more heterogeneous streamflow correlation, whereas streamflow dynamics are generally more correlated during spring and winter. Here equation (2) with streamflow-estimated parameters is employed. The grey circles refer to model performances during the other seasons. The lower boxplots highlight how observed and modeled streamflow correlation changes between seasons. The model properly reproduces the observed variability in streamflow correlation with some underestimations in summer and winter.

8 shows that the frequency and intensity of effective rainfall events are the main driver of streamflow spatial correlation, with a limited impact of heterogeneous drainage rates in most cases ($\langle F_k \rangle$=0.98). However, the left tail of the frequency distributions of F_k indicate that, in a limited number of cases, remarkable drops of

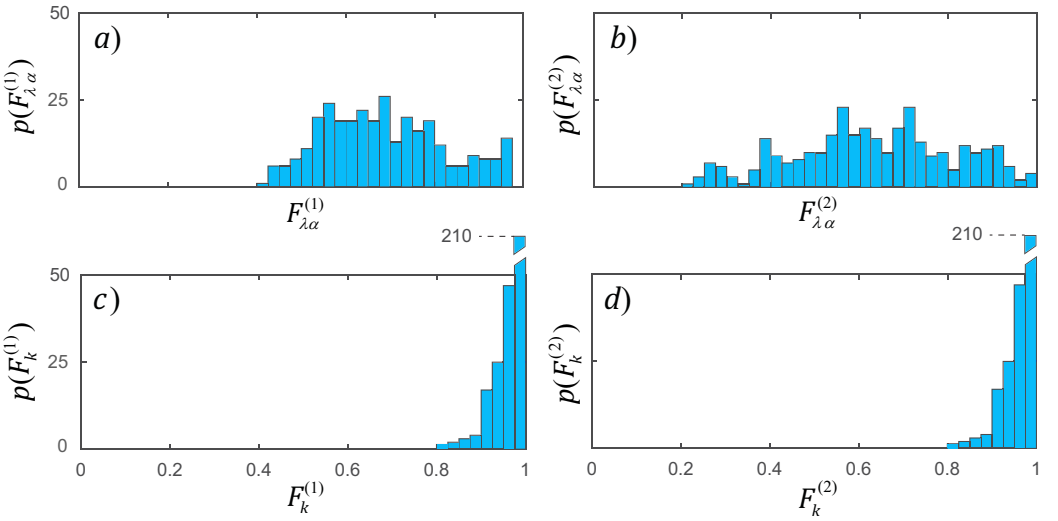

Figure 8. Frequency distributions of $F_{\lambda\alpha}^{(1)}$ and $F_k^{(1)}$ (equation (1)) and $F_{\lambda\alpha}^{(2)}$ and $F_k^{(2)}$ (equation (2)) for the 312 estimates of seasonal streamflow correlations. Values of F_* close to 1 mean that the hydrological process represented by $*$ has a limited impact on the streamflow correlation. Vice versa lower values of F_* identify the processes that have a stronger influence in reducing the correlation of daily flows between pairs of outlets. The corresponding histograms for equation (3) are similar to those of equation (2) and hence they are not shown here.

streamflow correlation can be induced by enhanced differences between the recession rates in the two catchments.

Figure 9 investigates how the heterogeneity of catchment-scale hydrological drivers (left plots), propagates to the streamflow correlation through the corresponding factors F_λ, F_α, and F_k (right plots). The analysis is carried out using equation (1) whose parameters are evaluated as detailed in section 4.2. To quantify the diversity of the relevant hydrological drivers between the two sites, a synthetic dimensionless index of heterogeneity is defined as $V(*)=\frac{|*_1-*_2|}{*_1+*_2}$ (where * can refer to λ_t, α^{12} or k) (Betterle et al., 2017). Since $V(*)\in[0, 1]$, $V(*)=0$ represents the case of perfect homogeneity, whereas $V(*)=1$ represents maximum heterogeneity between catchment pairs with respect to the parameter $*$. The frequency distributions of $V(\lambda_t)$, $V(\alpha^{12})$, and $V(k)$ are shown in the left plots of Figure 9. Additionally, the frequency distributions of r_α and $\frac{\lambda_{12}}{\lambda_m}$ ($\lambda_m=\min[\lambda_{1t}, \lambda_{12t}]$) are shown. In the left plots of Figure 9, the position of the histograms indicates the degree of spatial heterogeneity of the model parameters. The more each frequency distribution lies on the right, the higher the difference of the corresponding attribute between catchment pairs. Figure 9e shows that recession rates span a wide range of intercatchment heterogeneity. However, because of the reduced sensitivity of ρ_{model} on $V(k)$, the impact of intercatchment variability of recession rates on the correlation of daily flows is quite limited (Figure 9f). On the contrary, the frequencies of effective rainfall events are relatively uniform in different sites, and the relative fraction of joint events ($\frac{\lambda_{12}}{\lambda_m}$) exceeds 0.75 in most cases (Figure 9a). However, the impact of the heterogeneity of rainfall frequencies on the streamflow correlation is significant (Figure 9b) because of the high sensitivity of ρ_{model} to $V(\lambda_t)$ and $\frac{\lambda_{12}}{\lambda_m}$. The heterogeneity in the intensity of effective rainfall events, both in terms of spatial variability of the mean depths and in terms of lack of correlation between the joint depths, is quite pronounced (Figure 9c), leading to a notable loss of streamflow correlation, with $F_\alpha^{(1)} < 0.75$ in a significant number of cases. The marked sensitivity of streamflow correlation to the frequency and depth of joint flow-producing events demonstrates the key role of effective rainfall events in controlling spatial patterns of streamflow dynamics.

5.3. Estimation of Streamflow Regimes in Ungauged Catchments

Flow regimes represent the temporal variability of discharge during a specific period. The variability of flow conditions at a station is effectively embedded in the probability distribution of streamflows (PDF) or in the corresponding flow duration curve (FDC). Therefore, seasonal and annual streamflow PDFs and FDCs provide important indications for optimal management of water resources, risk assessment, ecological studies,

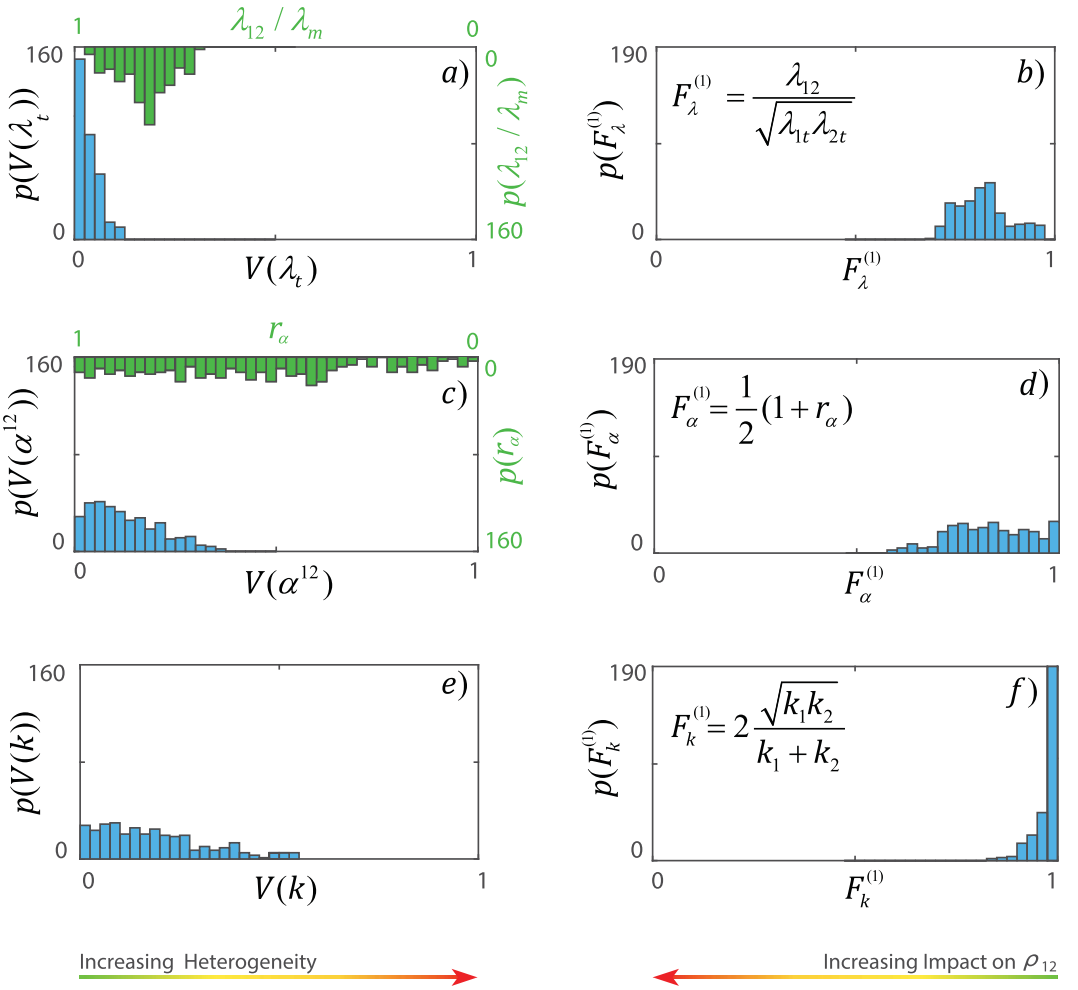

Figure 9. Frequency distributions of the variability index $V(*)=\frac{|*_1-*_2|}{*_1+*_2}$ ($*=\lambda_t,\ \alpha^{12},\ k$) and of the factors $F_\lambda^{(1)}$, $F_\alpha^{(1)}$ and $F_k^{(1)}$ for the 312 estimated seasonal streamflow correlations. Equation (1) with parameters estimated based on rainfall records is considered here. The histograms show how the factors F_*—and therefore streamflow correlation—have a different sensitivity to intercatchment heterogeneity in the corresponding hydrological descriptor $*$.

and river restoration practices. From an engineering perspective, for example, flow duration curves can provide valuable information for sizing artificial impoundments or in the design of run-of-the-river hydropower plants (Basso & Botter, 2012; Gorla & Perona, 2013; Lazzaro et al., 2013; Perona et al., 2013).

Streamflow PDF can be directly estimated from empirical frequency distributions (Castellarin et al., 2004, 2007; Vogel & Fennessey, 1995). Unfortunately, streamflow gauging stations are often lacking, they are unevenly distributed along river networks and/or available records might be too short for statistical inferences. Therefore, the estimation of flow PDFs at sites where no stream gauges are available (ungauged catchments (Blöschl et al., 2013)) is a key issue, with important scientific and practical consequences.

Here a new method to identify pairs of river sections characterized by similar flow regimes is presented. The method is suited to predict streamflow probability distributions in the absence of discharge data by taking advantage of the streamflow spatial correlations estimated by the analytical model employed in this paper. Our hypothesis is that flow regimes reflect the similarities of streamflow dynamics, as quantified by the spatial correlation of streamflows (Archfield & Vogel, 2010). Therefore, streamflow correlation could be efficiently used to identify river sites having analogous flow regimes.

Equation (1) with rainfall-estimated parameters (see section 4.2) is used to individuate, among the study catchments, those pairs that have high seasonal streamflow correlation and, therefore, are expected to share the same type of flow regime. Figure 10 shows the comparison between the normalized seasonal

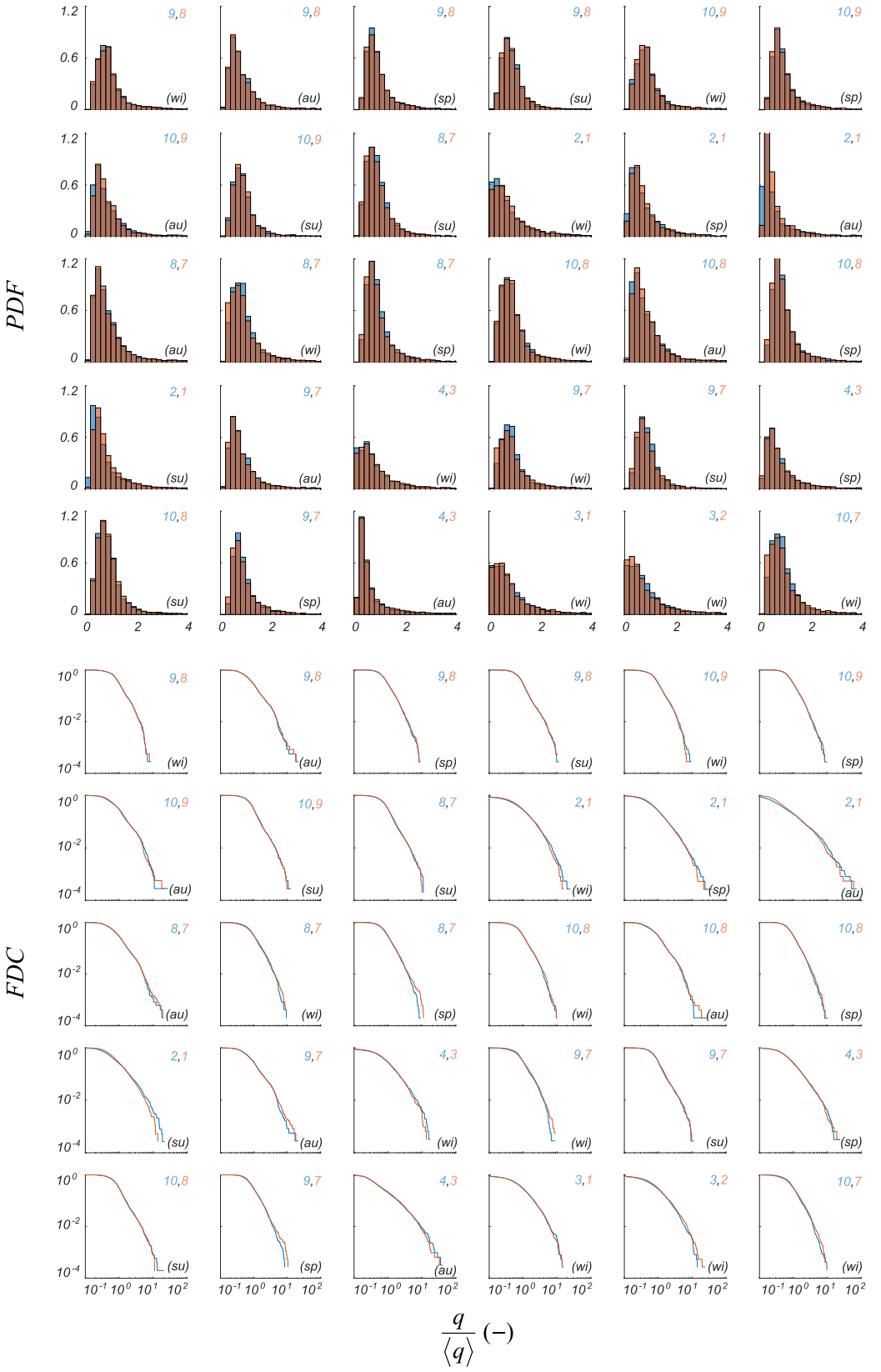

Figure 10. Comparison between the normalized seasonal streamflow PDF and FDC observed at all catchment outlets expected to have high streamflow correlation ($\rho^{(1)}_{model} > 0.9$). The couple of catchments considered in each plot is displayed (the numbers refer to catchments as in Table 2 and Figure 1) as well as the corresponding season (in brackets). Highly correlated catchments have very similar flow statistics.

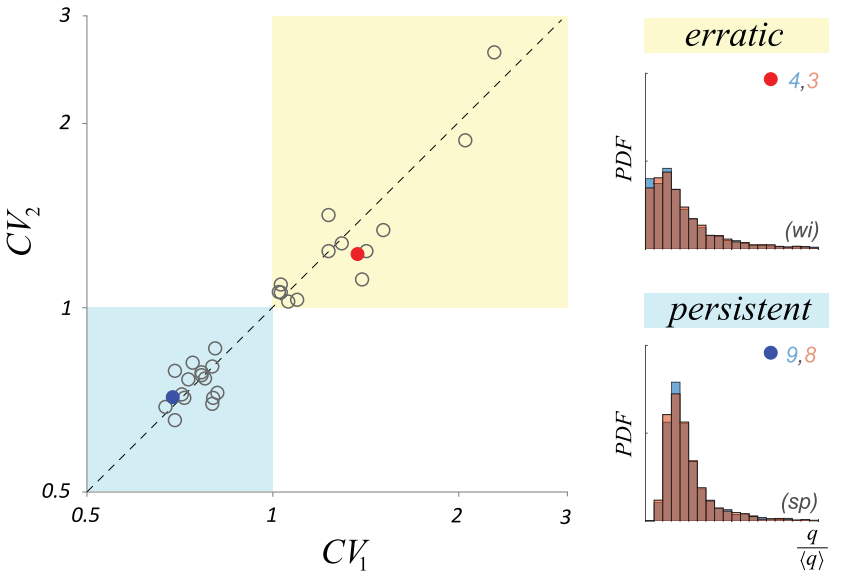

Figure 11. Comparison between the seasonal streamflow coefficient of variation (*CV*) between the pairs of outlets with $\rho^{(1)}_{model} > 0.9$. The framework is able to properly distinguish between erratic ($CV > 1$) and persistent ($CV < 1$) flow regimes (see text).

streamflow PDFs and FDCs observed at the outlets of all pairs for which $\rho^{(1)}_{model} > 0.9$. The plots clearly show how the model successfully identifies river sites characterized by very similar streamflow distributions. Note the different scales on the horizontal axis of the PDFs and FDCs plots, which highlights how similarities of flow statistics are not limited to the bulk of the distribution, but also include extreme events ($q \gg \langle q \rangle$).

The framework is appealing to regionalize streamflow PDFs since it can be used to group catchments with similar streamflow frequency distribution in the absence of discharge time series. Moreover, for any gauged location the method can easily be applied (without any calibration) to identify all ungauged sites with similar flow regimes, requiring only daily meteorological data readily available in most settings. The limited discrepancies between each couple of PDFs (FDCs) in Figure 10 witness the small errors that would be gathered by exporting the streamflow statistics from a donor gauged catchment to a receiver ungauged site within the study region using the proposed method.

In sparsely gauged areas, the framework can also be employed to classify streamflow regimes into erratic versus persistent based on their flow variability (see Botter et al., 2013). Erratic flow regimes characterize river sections that run dry quite often and whose flow dynamics display high variability (streamflow coefficient of variation $CV > 1$). On the other hand, persistent flow regimes are typical of river sites where flow is usually close to the mean and are characterized by lower coefficients of variation ($CV < 1$). Erratic flow regimes are normally associated to fast-responding catchments forced by sporadic effective rainfall. Instead, catchments that slowly release consistent amounts of water stored during frequent effective rainfall events are typically persistent (Botter et al., 2007c). Figure 11 shows the comparison between the *CVs* of catchment pairs for which $\rho^{(1)}_{model} > 0.9$. The plots suggest that this fundamental hydrological and ecological feature (namely the coefficient of variation of daily flows and the associated degree of erraticity/persistency of the regime) can be properly predicted by the proposed method, also in the absence of discharge data.

6. Conclusions

In this paper, a physically based analytical framework has been employed to investigate and quantify the drivers of the steady state correlation between daily synchronous streamflow time series at the outlet of two arbitrary catchments, at seasonal and annual timescale. To that aim, the model has been applied to a set of case studies located in a $75,000$ km^2 region in the eastern United States and used to evaluate the influence played by spatial heterogeneity of observed hydrological drivers on the resulting streamflow correlation. Additionally, a method to estimate model parameters in the absence of discharge data has been

proposed, which allows reasonable predictions of streamflow correlation in ungauged sites. The method is suitable to individuate catchment outlets characterized by similar flow regime and can be used to estimate streamflow frequency distributions in areas where discharge data are not available.

The following conclusions are worth emphasizing:

1. Model performances are satisfactory in most cases, with *RMSE* between observed and modeled correlations typically below 10%. This suggests that, in spite of the simplifications adopted, the main physical drivers of streamflow dynamics at pairs of outlets, and their influence on the spatial correlation of daily flows, are properly accounted for. Nonetheless, larger-scale studies and/or more extensive benchmarking will be necessary to better assess the merit and the potential of the formulation.
2. Equation (2) with parameters estimated from streamflow time series is the most accurate model in capturing the observed variability of streamflow correlation across the study catchments. Equation (3) with parameters estimated from discharge records provided performances similar to those of equation (2).
3. Equation (1) with parameters estimated directly from rainfall time series provided satisfactory results in reproducing the observed spatial variability of streamflow correlation. This method does not require any calibration on observed streamflow data; therefore, it is the best candidate to predict streamflow spatial correlations in settings where precipitation records are available, but hydrometric stations are lacking.
4. Correlation exhibits significant seasonal variability. On average, higher correlations are observed in winter and spring, whereas lower correlations and higher intercatchment variability are observed in autumn and summer. The relationship between correlation and distance also varies throughout the year, mainly in response to changes in the spatial structure of rainfall, showing longer correlation ranges in spring and winter.
5. Frequency and intensity of effective rainfall events are the main driver of daily streamflow correlation. This is a by-product of the pronounced sensitivity of the spatial correlation of discharge to heterogeneity in key properties of flow-producing rainfall events (especially the frequency and intensity of joint events).
6. In spite of the considerable variability of recession properties across the study catchments, heterogeneity in the drainage rates of the catchments bears in most cases a limited influence on the observed streamflow correlation.
7. The streamflow spatial correlation, predicted in the absence of discharge data, can be used to identify river sites characterized by similar flow regimes. Therefore, the proposed method can be employed to regionalize streamflow statistics by exporting streamflow PDFs and FDCs from gauged to ungauged sites.

In the proposed framework, key model parameters can be estimated from spatially averaged rainfall fields. Therefore, the model can be applied to any arbitrary site along a river network without requiring spatially distributed streamflow data or ad-hoc calibrations. As the approach accounts for the topological arrangement of catchments, it could help the design of spatially optimized discharge gauging networks and the redaction of streamflow correlation maps. Given its ability to quantify the influence of the heterogeneity of hydrological variables on flow characteristics and their spatial patterns, the methodology can assist studies concerning chemical, biological, and physical processes that are significantly impacted by the spatiotemporal variability or river flows and their underlying hydroclimatic drivers.

Acknowledgments
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement 641939. This study was also funded by the Swiss National Science Foundation (SNF, project 200021-149126). Additional support was provided by the Competence Center Environment and Sustainability (CCES) of the ETH domain in the framework of the RECORD Catchment project. All the hydrologic data used in this study are taken from the MOPEX data set and are freely available online at: http://www.nws.noaa.gov/ohd/mopex/. We thank the editors, two anonymous reviewers, and Marc Müller for their helpful comments.

References

Archfield, S. A., & Vogel, R. M. (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. *Water Resources Research*, *46*, W10513. https://doi.org/10.1029/2009WR008481

Basso, S., & Botter, G. (2012). Streamflow variability and optimal capacity of run-of-river hydropower plants. *Water Resources Research*, *46*, W10527. https://doi.org/10.1029/2012WR012017

Basso, S., Frascati, A., Marani, M., Schirmer, M., & Botter, G. (2015). Climatic and landscape controls on effective discharge. *Geophysical Research Letters*, *42*, 8441–8447. https://doi.org/10.1002/2015GL066014

Betterle, A., Schirmer, M., & Botter, G. (2017). Characterizing the spatial correlation of daily streamflows. *Water Resources Research*, *53*, 1646–1663. https://doi.org/10.1002/2016WR019195

Biswal, B., & Marani, M. (2014). Universal' recession curves and their geomorphological interpretation. *Advances in Water Resources*, *65*, 34–42. https://doi.org/10.1016/j.advwatres.2014.01.004

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (2013). *Runoff prediction in ungauged basins: Synthesis across processes, places and scales*. Cambridge, UK: Cambridge University Press.

Botter, G., Basso, S., Rodriguez-Iturbe, I., & Rinaldo, A. (2013). Resilience of river flow regimes. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(32), 12925–12930. https://doi.org/10.1073/pnas.1311920110

Botter, G., Peratoner, F., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2007c). Signatures of large-scale soil moisture dynamics on streamflow statistics across U.S. climate regimes. *Water Resources Research*, *43*, W11413. https://doi.org/10.1029/2007WR006162

Botter, G., Porporato, A., Daly, E., Rodriguez-Iturbe, I., & Rinaldo, A. (2007b). Probabilistic characterization of base flows in river basins: Roles of soil, vegetation, and geomorphology. *Water Resources Research*, *43*, W06404. https://doi.org/10.1029/2006WR005397

Botter, G., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2007a). Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. *Water Resources Research*, *43*, W02417. https://doi.org/10.1029/2006WR005043

Castellarin, A., Camorani, G., & Brath, A. (2007). Predicting annual and long-term flow-duration curves in ungauged basins. *Advances in Water Resources*, *30*(4), 937–953. https://doi.org/10.1016/j.advwatres.2006.08.006

Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., & Brath, A. (2004). Regional flow-duration curves: Reliability for ungauged basins. *Advances in Water Resources*, *27*(10), 953–965. https://doi.org/10.1016/j.advwatres.2004.08.005

Castiglioni, S., Castellarin, A., Montanari, A., Skøjen, J. O., Laaha, G., & Blöschl, G. (2011). Smooth regional estimation of low-flow indices: Physiographical space based interpolation and top-kriging. *Hydrology and Earth System Sciences*, *15*, 715–727. https://doi.org/10.5194/hess-15-715-2011

Castiglioni, S., Lombardi, I., Toth, E., Castellarin, A., & Montanari, A. (2010). Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach. *Advances in Water Resources*, *33*(10), 1235–1242. https://doi.org/10.1016/j.advwatres.2010.04.009

Ceola, S., Botter, G., Bertuzzo, E., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2010). Comparative study of ecohydrological streamflow probability distributions. *Water Resources Research*, *46*, W09502. https://doi.org/10.1029/2010WR009102

Chacon-Hurtado, J., Alfonso, L., & Solomatine, D. (2017). Rainfall and streamflow sensor network design: A review of applications, classification, and a proposed framework. *Hydrology and Earth System Sciences*, *21*, 3071–3091. https://doi.org/10.5194/hess-21-3071-2017

Cheng, B., & Krajewski, W. (2016). Analysing individual recession events: Sensitivity of parameter determination to the calculation procedure. *Hydrological Sciences Journal*, *61*(16), 2887–2901. https://doi.org/10.1080/02626667.2016.1170940

Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., & Sivapalan, M. (2012). Exploring the physical controls of regional patterns of flow duration curves Part 1: Insights from statistical analyses. *Hydrology and Earth System Sciences*, *16*, 4435–4446. https://doi.org/10.5194/hess-16-4435-2012

Claps, P., Giordano, A. & Laio, F. (2005). Advances in shot noise modeling of daily streamflows. *Advances in Water Resources*, *28*(9), 992–1000. https://doi.org/10.1016/j.advwatres.2005.03.008

Daly, C., Halbleib, M., Smith, J., Gibson, W., Doggett, M., Taylor, G.,. . . Pasteris, P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*, 2031–2064. https://doi.org/10.1002/joc.1688

Doulatyari, B., Betterle, A., Basso, S., Biswal, B., Schirmer, M., & Botter, G. (2015). Predicting streamflow distributions and flow duration curves from landscape and climate. *Advances in Water Resources*, *83*, 285–298. https://doi.org/10.1016/j.advwatres.2015.06.013

Doulatyari, B., Betterle, A., Radny, D., Alessi Celegon, E., Fanton, P., Schirmer, M., & Botter, G. (2017). Patterns of streamflow regimes along the river network: The case of the Thur River. *Environmental Modelling and Software*, *93*, 42–58.

Dralle, D., Karst, N., & Thompson, S. (2015). a, b careful: The challenge of scale invariance for comparative analyses in power law models of the streamflow recession. *Geophysical Research Letters*, *42*, 9285–9293. https://doi.org/10.1002/2015GL066007

Gorla, L., & Perona, P. (2013). On quantifying ecologically sustainable flow releases in a diverted river reach. *Journal of Hydrology*, *489*, 98–107. https://doi.org/10.1016/j.jhydrol.2013.02.043

Hansen, M., De Fries, R., Townshend, J., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree. *International Journal of Remote Sensing*, *21*(6–7), 1331–1364. https://doi.org/10.1080/014311600210209

Isaak, D. J., Peterson, E. E., Ver Hoef, J. M., Wenger, S. J., Falke, J. A., Torgersen, C. E.,. . . Monestiez, P. (2014). Applications of spatial statistical network models to stream data. *WIREs Water Advanced Review*, *1*, 277–294. https://doi.org/10.1002/wat2.1023

Jefferson, A., Grant, G., Lewis, S., & Lancaster, S. (2010). Coevolution of hydrology and topography on a basalt landscape in the Oregon Cascade Range, USA. *Earth Surface Processes and Landforms*, *35*, 803–816. https://doi.org/10.1002/esp.1976

Huang, X., & Niemann, J. (2006). Modelling the potential impacts of groundwater hydrology on long-term drainage basin evolution. *Earth Surface Processes and Landforms*, *31*, 1802–1823. https://doi.org/10.1002/esp.1369

Lai, C.-T., & Katul, G. (2000). The dynamic role of root-water uptake in coupling potential to actual transpiration. *Advances in Water Resources*, *23*(4), 427–439. https://doi.org/10.1016/S0309-1708(99)00023-8

Laio, F., Porporato, A., Ridolfi, L., & Rodriguez-Iturbe, I. (2001). Plants in water-controlled ecosystems: Active role in hydrologic processes and response to water stress, II. Probabilistic soil moisture dynamics. *Advances in Water Resources*, *24*(7), 707–723. https://doi.org/10.1016/S0309-1708(01)00005-7

Lahaa, G., Skøjen, J. O., & Blöschl, G. (2014). Spatial prediction on river networks: Comparison of top-kriging with regional regression. *Hydrological Processes*, *28*, 315–324. https://doi.org/10.1002/hyp.9578

Lazzaro, G., Basso, S., Schirmer, M., & Botter, G. (2013). Water management strategies for run-of-river power plants: Profitability and hydrologic impact between the intake and the outflow. *Water Resources Research*, *49*, 8285–8298. https://doi.org/10.1002/2013WR014210

Mc Guire, K., Torgersen, C., Likens, G., Buso, D., Lowe, W., & Bailey, S. (2014). Network analysis reveals multiscale controls on streamwater chemistry. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(19), 7030–7035. https://doi.org/10.1073/pnas.1404820111

Mejia, A., Daly, E., Rossel, F., Jovanovic, T., & Gironas, J. (2014). A stochastic model of streamflow for urbanized basins. *Water Resources Research*, *50*, 1984–2001. https://doi.org/10.1002/2013WR014834

Messinger, T., & Paybins, K. (2014). Correlations of daily flows at streamgages in and near West Virginia, 19302011, and streamflow characteristics relevant to the use of index streamgages. *U.S. Geological Survey, Scientific Investigations* (Report 2014–5061). Reston, VA: U.S. Geological Survey.

Milly, P. C. D. (1994). Climate, soil water storage, and the average annual water balance. *Water Resources Research*, *30*(7), 2143–2156. https://doi.org/10.1029/94WR00586

Müller, M. F., Dralle, D. N., & Thompson, S. E. (2014). Analytical model for flow duration curves in seasonally dry climates. *Water Resources Research*, *50*, 5510–5531. https://doi.org/10.1002/2014WR015301

Müller, M. F., & Thompson, S. E. (2015). TopREML: A topological restricted maximum likelihood approach to regionalize trended runoff signatures in stream networks. *Hydrology and Earth System Sciences*, *19*, 2925–2942. https://doi.org/10.5194/hess-19-2925-2015

Perona, P., Dürrenmatt, D., & Characklis, G. (2013). Obtaining natural-like flow releases in diverted river reaches from simple riparian benefit economic models. *Journal of Environmental Management*, *118*, 161–169. https://doi.org/10.1016/j.jenvman.2013.01.010

Porporato, A., Daly, E., & Rodriguez-Iturbe, I. (2004). Soil water balance and ecosystem response to climate change. *The American Naturalist*, *164*(5), 625–632. https://doi.org/10.1086/424970

Postel, S., & Richter, B. (2003). *Rivers for life: Managing water for people and nature*. Washington, DC: Island Press. https://doi.org/10.1002/rra.820

Pumo, D., Noto, L. V., & Viola, F. (2013). Ecohydrological modelling of flow duration curve in Mediterranean river basins. *Advances in Water Resources*, *52*, 314–327. https://doi.org/10.1016/j.advwatres.2012.05.010

Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V., & Coxi, D. R. (1999). Probabilistic modelling of water balance at a point: The role of climate, soil and vegetation. *Proceedings of the Royal Society*, *455*, 3789–3805. https://doi.org/10.1098/rspa.1999.0477

Sabo, J. L., Sinha, T., Bowling, L. C., Schoups, G. H. W., Wallender, W. W., Campana, M. E.,. . . Wohln, E. E. (2010). Reclaiming sustainable watersheds in the Cadillac Desert. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21263–21269. https://doi.org/10.1073/pnas.1009734108

Schaake, J., Cong, S., & Duan, Q. (2006). *U.S. MOPEX data set* (IAHS publication series, pp. 1–23).

Schaefli, B., Nicotina, L., Imfeld, C., Da Ronco, P. X., Bertuzzo, E., & Rinaldo, A. (2014). Spatially explicit hydrologic response model for ecohydrologic applications. *Geoscientific Model Development*, *7*, 2733–2746. https://doi.org/10.5194/gmd-7-2733-2014

Settin, T., Botter, G., Rodriguez-Iturbe, I., & Rinaldo, A. (2007). Numerical studies on soil moisture distributions in heterogeneous catchments. *Water Resources Research*, *43*, W05425. https://doi.org/10.1029/2006WR005737

Sivapalan, M., Thompson, S. E., Harman, C. J., Basu, N. B., & Kumar, P. (2011). Water cycle dynamics in a changing environment: Improving predictability through synthesis. *Water Resources Research*, *47*, W00J01. https://doi.org/10.1029/2011WR011377

Skøjen, J. O., & Blöschl, G. (2007). Spatiotemporal topological kriging of runoff time series. *Water Resources Research*, *43*, W09419. https://doi.org/10.1029/2006WR005760

Skøjen, J. O., Merz, R., & Blöschl, G. (2006). Top-kriging—Geostatistics on stream networks. *Hydrology and Earth System Sciences*, *10*, 277–287. https://doi.org/10.5194/hess-10-277-2006

Thompson, S. E., Harman, C. J., Konings, A. G., Sivapalan, M., Neal, A., & Troch, P. A. (2011). Comparative hydrology across AmeriFlux sites: The variable roles of climate, vegetation, and groundwater. *Water Resources Research*, *47*, W00J07. https://doi.org/10.1029/2010WR009797

Verma, P., Yeates, J., & Daly, E. (2011). A stochastic model describing the impact of daily rainfall depth distribution on the soil water balance. *Advances in Water Resources*, *34*(8), 1039–1048. https://doi.org/10.1016/j.advwatres.2011.05.013

Vogel, R. M., & Fennessey, N. M. (1995). Flow duration curves II. A review of application in water-resources planning. *Water Resources Bulletin*, *31*(6), 1029–1039.

Widder, S., Besemer, K., Singer, G. A., Ceola, S., Bertuzzo, E., Quince, C.,. . . Battin, T. J. (2014). Fluvial network organization imprints on microbial co-occurrence networks. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(35), 12799–12804. https://doi.org/10.1073/pnas.1411723111