UNIVERSITÉ DE NEUCHÂTEL

# PREDICTING THE SPATIAL CORRELATION OF DAILY STREAMFLOWS TO UNDERSTAND THE HYDROLOGICAL RESPONSE OF RIVER NETWORKS

Thesis presented to the Science Faculty of the University of Neuchâtel to satisfy the requirements of Philosophy Doctor in Science by:

ANDREA BETTERLE

Supervisory Commitee:

Prof. Mario Schirmer, University of Neuchâtel (Director of the thesis)
Prof. Gianluca Botter, University of Padova (Technical supervisor)
Prof. Paolo Perona, University of Edinburgh (External Reviewer)
Prof. Philippe Renard, University of Neuchâtel (Reviewer)

Université de Neuchâtel

24 October 2018

UNIVERSITÉ DE NEUCHÂTEL

**Faculté des Sciences**
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

# IMPRIMATUR POUR THESE DE DOCTORAT

**La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par**

## Monsieur Andrea BETTERLE

Titre:

## “Predicting the spatial correlation of daily streamflows to understand the hydrological response of river networks”

**sur le rapport des membres du jury composé comme suit:**

- Prof. associé Mario Schirmer, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. associé, Philippe Renard, Université de Neuchâtel, Suisse
- Prof. associé Gianluca Botter, Università degli Studi di Padova, Italia
- Prof. Paolo Perona, University of Edinburgh, United Kingdom

Neuchâtel, le 10 décembre 2018

Le Doyen, Prof. P. Felber

In distant antiquity, at times the rains would come, at times they would not. No one knew when. At times it would be cold, at times it would be hot. When it was cold, humans, who had no clothes to wear, were at risk of freezing to death. When rain did not fall, plants did not grow. When the rains came, plants and berries grew, which humans could eat to nourish themselves, but just as often the plants were poison, and made them ill. Gradually humans began to understand that these events were not random. They came to realize when it would rain and when it would not; when it would be cold and when it would be warm. They began to realize which plants they could eat, and which were poison. They began to domesticate the plants. They would plant them according to the changes in the weather, which they came to know as the seasons. Eventually, what had once seemed like unpredictable chaos of natural phenomena—random rains, wind, cold, heat, nourishment, and poison—were turned into a harmonious system. That which grew from the earth was now correlated with the larger patterns of the heavens. But this was not natural. Humans had domesticated the world. Humans had made it so that these disparate phenomena became a harmonious set of processes.

— Xunzi

# SUMMARY

Rivers are fascinating expressions of the hydrological cycle and their waters, riparian areas and floodplains harbor a multitude of ecosystems. The potential for rivers to sustain life depends on the dynamics of water quantity – and quality – that are critically controlled by catchment-scale geomorphoclimatic processes. In particular, the temporal and spatial availability of streamflows is paramount for rivers to provide ecosystem service and for the anthropic use of water resources. Therefore, understanding the hydroclimatic controls of streamflow dynamics is crucial to predict discharge dynamics at arbitrary locations and to cope with extreme events like floods and droughts. Moreover, revealing the links between the spatiotemporal variability of geomorphoclimatic traits and the ensuing patterns of streamflows dynamics can help disclosing hydrological legacies on the ecological functioning of river networks.

Understanding the dynamic of river flows is challenging because of the complexity, non-linearity and feedback-like nature of many hydrological processes. Advances in the mechanistic interpretation of the water cycle are thus necessary to support predictions of the spatial and temporal variability of streamflows. In particular, enhanced descriptions of flow dynamics can provide the basis to tackle in a more conscious way the wide array of practical and scientific challenges of our times. Optimal management of water resources (e.g. hydropower production), flood and droughts protection (via the management of flood detention basins and early warning systems), or stream ecology (though river restoration), are some of the contexts where predictions of hydrological fluxes are critical.

This thesis focuses on the spatial correlation of river flows, namely the correlation between daily streamflow timeseries at two generic locations along a river network. The correlation between river flows is found to be a powerful indicator that quantifies the similarity between a range of features characterizing the hydrological response of two catchments. As a statistical index, correlation can be considered as a metric quantifying the synchronicity of two signals. However, we find that average seasonal flows, streamflow variability, as well as other features defining the hydrological response of catchments, are extremely similar at sites displaying highly correlated flows. These evidences suggest a deep relationship between the physical processes that control the spatial correlation of river flows and those responsible for the hydrological response of the landscape.

A physically-based stochastic model was developed to quantify the correlation between daily flows at two river sites. The model is

based on a simple description of main catchment-scale hydrological processes and it requires basic hydroclimatic inputs. Despite its parsimony, the model succeeds in reproducing observed streamflow correlations in absence of discharge data, without requiring calibrations. Model predictions of streamflow correlation are useful in cases of limited hydrological information to export streamflow timeseries and other flow statistics from gauged to ungauged locations. Especially where direct flow measurements are not available, the model can be used to identify hydrologically similar locations, providing a means to spatially extend point information on flow regimes. The model was employed to assess how the different physical processes underlying flow dynamics ultimately affect streamflow correlation. The approach enables a proper quantification of the effect due to inter-catchment heterogeneities of runoff-driving processes on the ensuing hydrological similarity of two river basins. Spatial correlation is primarily controlled by the frequency and the intensity of rainfall events that simultaneously generate runoff in both catchments. The larger the number of rainfall events that simultaneously generate runoff in both catchments compared to the total set of runoff-generating events, the more similar will be the hydrological response of the two sites. Moreover, when simultaneous effective rainfall events have similar intensities, streamflow correlation will further increases. Interestingly, inter-catchment differences in response rates – namely how quickly a catchment drains after a rainfall event – mildly affect the correlation between daily flows.

The work presented in this thesis offers a new interpretation on how spatially heterogeneous geomorphoclimatic drivers of the water cycles ultimately affect seasonal patterns of flow regimes along river networks. In an era of unsustainable anthropogenic pressures on the environment, the novel insights provided by this study offer new perspectives to understand and better manage hydrologic systems and river networks.

Keywords:

Streamflow correlation; Spatial variability; Streamflow dynamics; Climate; Morphology; Hydrology; River flow regimes; Stochastic analytical model; Flow patterns; Runoff prediction in ungauged basins (PUB); Streamflow frequency distributions; Flow duration curves; Average seasonal flow; Streamflow recession; Catchment response rate; Hydromorphology; Ecohydrology; Physically-based, Data scarcity.

# RÉSUMÉ

Les rivières sont des expressions fascinantes du cycle hydrologique, et leurs eaux, les zones riveraines et les plaines inondables abritent une multitude d'écosystèmes. Le potentiel des rivières de soutenir la vie dépend de la dynamique de la quantité, et de la qualité, de l'eau, qui sont contrôlées de façon critique par des processus géomorphoclimatiques à l'échelle du bassin versant. En particulier, la disponibilité des débits, dans le temps et l'espace, des cours d'eau est primordiale pour que les rivières fournissent un service écosystémique et pour l'utilisation anthropique des ressources en eau. Par conséquent, il est essentiel de comprendre les contrôles hydroclimatiques sur la dynamique du débit des cours d'eau pour prédire la dynamique du débit à des endroits arbitraires, et pour faire face aux événements extrêmes comme les inondations et les sécheresses. De plus, la mise en évidence des liens entre la variabilité spatio-temporelle des caractéristiques géomorphoclimatiques et les schémas qui en découlent dans la dynamique des débits fluviaux peut aider à révéler les héritages hydrologiques sur le fonctionnement écologique des réseaux hydrographiques.

Il est difficile de comprendre la dynamique des débits fluviaux en raison de la complexité, de la non-linéarité et de la nature rétroactive de nombreux processus hydrologiques. Des progrès dans l'interprétation mécaniste du cycle de l'eau sont donc nécessaires pour appuyer les prévisions de la variabilité spatiale et temporelle des débits des cours d'eau. En particulier, des descriptions améliorées de la dynamique des débits peuvent servir de base pour s'attaquer de manière plus consciente au large éventail de défis pratiques et scientifiques de notre époque. La gestion optimale des ressources en eau , la protection contre les inondations et les sécheresses ou l'écologie des cours d'eau sont quelques-uns des contextes où les prévisions des flux hydrologiques sont essentielles.

Cette thèse porte sur la corrélation spatiale des débits fluviaux, notamment la corrélation entre les séries chronologiques quotidiennes des débits à deux endroits génériques sur le long d'un réseau fluvial. La corrélation entre les débits fluviaux s'avère être un indicateur puissant qui quantifie la similitude entre une série de traits caractérisant la réponse hydrologique de deux bassins versants. En tant qu'indice statistique, la corrélation peut être considérée comme une métrique quantifiant la synchronicité de deux signaux. Cependant, nous constatons que les débits saisonniers moyens, la variabilité du débit, ainsi que d'autres caractéristiques définissant la réponse hydrologique des bassins versants, sont extrêmement similaires aux sites présentant

des débits fortement corrélés. Ces preuves suggèrent une relation profonde entre les processus physiques qui contrôlent la corrélation spatiale des débits fluviaux et ceux responsables de la réponse hydrologique du paysage.

Un modèle stochastique basé sur la physique a été développé pour quantifier la corrélation entre les débits quotidiens à deux sites fluviaux. Le modèle est basé sur une description simple des principaux processus hydrologiques à l'échelle du bassin versant et il nécessite des apports hydroclimatiques de base. Malgré sa parcimonie, le modèle réussit à reproduire les corrélations de débit observées en l'absence de données sur les débits, sans nécessiter d'étalonnage. Les prédictions modélisées de la corrélation des débits sont utiles dans les cas où l'information hydrologique est limitée, et permettent d'exporter des séries chronologiques et d'autres statistiques sur les débits de cours d'eau d'un endroit jaugé vers un autre non jaugé. En particulier lorsque des mesures directes de débit ne sont pas disponibles, le modèle peut être utilisé pour identifier des sites hydrologiquement similaires, ce qui permet d'étendre spatialement les informations ponctuelles sur les régimes des débits. Le modèle a été utilisé pour évaluer comment les différents processus physiques qui sous-tendent la dynamique des débits influent en fin de compte sur la corrélation des débits. Cette approche permet de quantifier correctement l'effet de l'hétérogénéité entre les bassins versants sur les processus de ruissellement sur la similarité hydrologique de deux bassins hydrographiques qui en résulte. La corrélation spatiale est principalement contrôlée par la fréquence et l'intensité des précipitations qui génèrent simultanément le ruissellement dans les deux bassins versants. Plus le nombre d'événements pluvieux qui génèrent simultanément du ruissellement dans les deux bassins versants est élevé par rapport au nombre total d'événements générateurs de ruissellement, plus la réponse hydrologique des deux sites sera similaire. De plus, lorsque des précipitations efficaces simultanées ont des intensités similaires, la corrélation des débits augmente encore. Il est intéressant de noter que les différences entre les bassins versants dans les taux de réponse - à savoir la vitesse à laquelle un bassin versant se draine après une pluie - affectent légèrement la corrélation entre les débits quotidiens.

Les travaux présentés dans cette thèse offrent une nouvelle interprétation de la façon dont les facteurs géomorphoclimatiques spatialement hétérogènes des cycles de l'eau influent en fin de compte sur les régimes d'écoulement saisonniers des réseaux fluviaux. À une époque où les pressions anthropiques sur l'environnement sont insoutenables, les nouvelles perspectives qu'offre cette étude offrent de nouvelles perspectives pour comprendre et mieux gérer les systèmes hydrologiques et les réseaux hydrographiques.

Mots-clés:

Corrélation des débits fluviaux; Variabilité spatiale; Dynamique des débits fluviaux; Climat; Morphologie; Hydrologie; Régimes des débits fluviaux; Modèle analytique stochastique; Modèles d'écoulement; Prévision du ruissellement dans les bassins non jaugés; Distribution des fréquences des débits; Courbe de durée; Débit saisonnier moyen; Recul du débit fluvial; Bassin hydrographique; Hydromorphologie; Ecohydrologie; Données physiques.

# PUBLICATIONS

Betterle A., Schirmer M., Botter G., (2017a), Characterizing the spatial correlation of daily streamows, *Water Resources Research*, https://doi.org/10.1002/2016WR019195

Betterle A., Radny D., Schirmer M., Botter G., (2017b), What Do They Have in Common? Drivers of Streamflow Spatial Correlation and Prediction of Flow Regimes in Ungauged Locations, *Water Resources Research*, https://doi.org/10.1002/2017WR021144

Betterle A., Schirmer M., Botter G., (2018), Flow dynamics at the continental scale: streamow correlation and hydrological similarity, *Hydrological Processes*, https://doi.org/10.1002/hyp.13350

# ACKNOWLEDGEMENTS

My gratitude goes to my PhD advisor Gianluca Botter for the professional, though friendly, support he provided me in these years and for being always available despite the 500 km distance and his packed schedule. Thanks to my supervisor Mario Schirmer for his comprehension and willingness to help me in any possible way. Thanks to my office mates (and ex office neighbours) Christian Möck, Reto Britt and Dirk Radny for the good (and bad) jokes of all these years. I would also like to tank my colleague Robin Weatherl and her "Frenchman" for translating in french the summary of this thesis. Last but not least I would like to thank my family, Valentina, Michele, Francesco and Annette for supporting me until now. A special thought goes to my grandparents Vittorio, Bruno, Antonietta and Vanilla. Although you are not with us any more, I would have not reached this goal if it wasn't for you. Zena, I always miss our time together.

# CONTENTS


Summary . . . . . v
Keywords . . . . . vii
Résumé . . . . . ix
Mots-clés . . . . . xi
Publications . . . . . xiii
Acknowledgements . . . . . xv
List of Figures xix
List of Tables xxviii
1 INTRODUCTION 1
1.1 Aims and objectives . . . . . 8
1.2 Thesis structure . . . . . 9
2 CHARACTERIZING THE SPATIAL CORRELATION OF DAILY STREAMFLOWS 11
2.1 Introduction . . . . . 12
2.2 Analytical characterization of streamflow correlation . 15
2.3 Exponentially distributed effective rainfall depths . . . 19
2.3.1 Case 1 . . . . . 21
2.3.2 Case 2 . . . . . 22
2.3.3 Case 3 . . . . . 23
2.4 Application . . . . . 24
2.5 Effect of spatially heterogeneous hydrological properties on the streamflow correlation . . . . . 26
2.6 Discussion . . . . . 30
2.7 Conclusions . . . . . 32
2.8 Appendix . . . . . 33
2.8.1 Analytical derivation of the streamflow correlation . . . . . 33
2.9 Effect of heterogeneous effective rainfall depths on $\rho_{12}$ 35
3 WHAT DO THEY HAVE IN COMMON? DRIVERS OF STREAMFLOW SPATIAL CORRELATION AND PREDICTION OF FLOW REGIMES IN UNGAUGED LOCATIONS 37
3.1 Introduction . . . . . 38
3.2 Methods . . . . . 40
3.3 Case studies and hydrologic data . . . . . 43
3.4 Parameter estimation . . . . . 45
3.4.1 Method A: Estimate of model parameters from streamflow time series . . . . . 45
3.4.2 Method B: Estimate of model parameters in the absence of discharge data . . . . . 46
3.5 Results and Discussion . . . . . 48
3.5.1 Prediction of streamflow spatial correlation and its seasonality . . . . . 48

3.5.2 Analysis of the impact of inter-catchment variability of hydroclimatic features on the streamflow spatial correlation . . . . . 54
3.5.3 Estimation of streamflow regimes in ungauged catchments . . . . . 58
3.6 Conclusions . . . . . 61
4 FLOW DYNAMICS AT THE CONTINENTAL SCALE: STREAMFLOW CORRELATION AND HYDROLOGICAL SIMILARITY 63
4.1 Introduction . . . . . 64
4.2 Methods . . . . . 68
4.3 Estimation of model parameters . . . . . 70
4.4 Study sites and hydrological data . . . . . 72
4.5 Performance metrics and indexes of similarity . . . . . 73
4.6 Continental-scale prediction of seasonal streamflow correlation . . . . . 75
4.7 Streamflow correlation and hydrological similarity . . 80
4.8 Prediction of the hydrologic response using correlation and distance . . . . . 85
4.9 Effect of streamgauge density on the selection of reference sites . . . . . 89
4.10 Discussion . . . . . 91
4.11 Conclusions . . . . . 94
5 CONCLUSIONS AND PERSPECTIVES 97
5.1 Conclusions . . . . . 97
5.2 Perspectives . . . . . 100

BIBLIOGRAPHY 105

# LIST OF FIGURES


Figure 1 Comparison between the streamflow timeseries at two pairs of catchment outlets in the Eastern United States. Figure 1a: Little Pigeon River at Sevierville (TN) ($q_1$) and Oconaluftee River at Birdtown (NC) ($q_2$), during autumn; Figure 1b: Valley River at Tomotla (NC) ($q_1$) and North Fork Holston River near Gate City (VA) ($q_2$), during spring. The insets show the scatter plots of $q_1$ and $q_2$. . . . . 14

Figure 2 When a pair of catchments are simultaneously considered, the overall timeseries of effective rainfall ($\xi_{it}$) within each catchment can be decoupled into two sub-timeseries. The first includes the joint effective rainfall events (when streamflow jumps are observed at the outlet of both catchments simultaneously) $\xi_i^{12}$, whereas the other includes the effective rainfall events experienced only by one of the two catchments, $\xi_i$. . . . . 17

Figure 3 The scatterplot shows the performances of the model by comparing the observed and the modeled daily streamflow correlation between all the possible couples of catchment outlets within the study region. The application is performed at seasonal timescale: a single dot represents two catchments during a given season. . . . . 26

Figure 4 The plots show how the three factors in the analytical formulation change as a function of the heterogeneity of the physical parameters involved. The parameters include: i) the average effective rainfall frequencies ($\lambda_{it}$, $\lambda_{12}$); ii) the average joint effective rainfall depths ($\alpha_i^{12}$) and the correlation $r_\alpha$ between the joint effective rainfall depths (expressed by $a = r_\alpha \frac{\alpha_2^{12}}{\alpha_1^{12}}$); iii) the streamflow recession rates ($k_i$). The derivative of each factor (right panels) highlights the sensitivity of $\rho_{12}$ to the heterogeneity of these parameters. In panels 4c and 4d it is assumed $\alpha_2^{12} > \alpha_1^{12}$. The stars in Figure 4b and 4d indicate step-changes in $F_\lambda$ and $F_\alpha$ (and the y coordinate of each star indicates the relative extent of the corresponding step change). . . . . 28

Figure 5 Implications of catchment arrangement within a river network for the flow correlation. Two configurations are possible: i) nested catchments: a smaller catchment ($A_1$) is nested within a larger one ($A_2$); ii) non nested catchments: the two catchment areas do not overlap. Due to downstream propagation of streamflows, in case of nested catchments, the frequency of joint runoff events equals the frequency of runoff events in the smaller catchment ($\lambda_{12} = \lambda_{1t}$). In case of non-nested catchments the frequency of joint events is usually smaller than the minimum runoff frequency in the two catchment ($\lambda_{12} < \min\{\lambda_{1t}, \lambda_{2t}\}$). . . . . 29

Figure 6 Effect of heterogeneous joint and disjoint effective rainfall intensities on streamflow correlation for different frequencies of effective rainfall. Differences between joint and disjoint effective rainfall intensities within the same catchment $i$ are quantified by the ratio $\delta_i = \frac{\alpha_i}{\alpha_i^{12}}$. The couple of ratios $(\delta_1, \delta_2)$ identify each curve. Without loss of generality it is assumed $\lambda_{1t} > \lambda_{2t}$. In panel a) the curves are independent on the parameter $\delta_2$ (because $\lambda_2 = 0$). The continuous red curve corresponds to equation (11) and Figure 4. As defined in the text, $\lambda_m = \min\{\lambda_{1t}, \lambda_{2t}\}$. . . . . 36

Figure 7 The study area includes 13 nested and non-nested catchments located in the eastern United States. . . . . 44

Figure 8 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 ($\lambda_t$, $\lambda_{12}$, $\lambda_i$). 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 ($r_\alpha$) and the correlation of the joint rainfall depths ($r_{\tilde{\alpha}}$) . . . . . 49

Figure 9 The scatterplots compare the seasonal streamflow correlation calculated from daily discharge records (x-axis) and the streamflow correlation estimated by the analytical model (equations (29) to (31)) when the parameters are obtained from streamflow records (y-axis) for the 312 pairs of seasonal streamflow time series between the 13 study catchments. Scatterlplots a), b) and c) refer to equations (29), (30) and (31) respectively. Scatterplot d) shows the performances of equation (30) when the intensities of joint effective rainfall events are estimated directly from rainfall records. . . . . 50

Figure 10 The scatterplots compare the streamflow correlation calculated from daily discharge records (x-axis) and the streamflow correlation estimated by the analytical model (equations (29) to (31)) 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 equation 29, 30 and 31 respectively assuming $F_k = 1$. Scatterplot d) shows the performances of equation (29) when the streamflow recession parameters are estimated based on discharge data. . . . . 51

Figure 11 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 panel a) streamflow correlation is estimated by means of equation (29) when $F_\lambda$ and $F_\alpha$ are obtained from rainfall data and $F_k$ is estimated from discharge time series. In panel b) streamflow correlation is estimated by means of equation (30) when $F_\lambda$ and $F_k$ are obtained from streamflow records and $F_\alpha$ is obtained from rainfall records. In panel c) streamflow correlation is estimated by means of equation (31) when model parameters are obtained from streamflow records. . . 52

Figure 12 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 (30) with parameters estimated from discharge as a function of the inter-catchment distance. . . . . 53

Figure 13 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 (30) 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. . . . . 55

Figure 14 Frequency distributions of $F^{(1)}_{\lambda\alpha}$ and $F^{(1)}_{k}$ (equation (29)) and $F^{(2)}_{\lambda\alpha}$ and $F^{(2)}_{k}$ (equation (30)) 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 (31) are similar to those of equation (30) and hence they are not shown here. . . . 56

Figure 15 Frequency distributions of the variability index $V(*) := \frac{|*_1 - *_2|}{*_1 + *_2}$ ($* = \lambda_t$, $\alpha^{12}$, $k$, and $1, 2$ identify the relevant catchments) and of the factors $F^{(1)}_{\lambda}$, $F^{(1)}_{\alpha}$ and $F^{(1)}_{k}$ for the 312 couples of estimated seasonal streamflow correlations. Equation (29) 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 inter-catchment heterogeneity in the corresponding hydrological descriptor $*$. . 57

Figure 16 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 Tab. 4 and Fig. 7) as well as the corresponding season (in brackets). Highly correlated catchments have very similar flow statistics. . . . . 59

Figure 17 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). . . . . 60

Figure 18 The 413 study catchments display a wide variety of geomorphoclimatic conditions. Inter catchment differences in terms of catchment sizes, shapes, topological arrangement (i.e. nested Vs non-nested catchments) and morphology suggest enhanced differences in hydrological responses. The maps show: a) Distribution and delineation of the study catchments; b) Seasonality (defined as $\sum_{s=1}^{4} |p_s/PET_s - p_a/PET_a|/(p_a/PET_a)$, where $p_s$ ($p_a$) and $PET_s$ ($PET_a$) are the average seasonal (and annual) precipitation and potential evapotranspiration depths; c) Elevation; d) Annual aridity (defined as $PET_a/p_a$); e) Land cover. The boxplots in panel f) summarizes the distribution of the main physical properties across the study sites. . . . . 74

Figure 19 Comparison between predicted and measured seasonal streamflow correlations at the outlets of each possible pair of the 413 study sites (340,312 pairs). Panel a): equation (37) with rainfall-estimated model parameters; panel b): equation (38) with discharge-estimate model parameters. . . . . 76

Figure 20 Distributions of the values assumed by the three factors that constitute the analytical expression for the streamflow correlation $\rho_{model}$ (equation (37)) across the study sites. Values close to 0 refer to cases where the inter-catchment heterogeneity in the physical processes represented by the corresponding factor F significantly contribute to a loss of streamflow correlation. . . . . 78

Figure 21 Intra-annual dynamics of runoff coefficients (showed in the boxplots) possibly affect the seasonal variability of streamflow correlation and model performances. Anticorrelated flows and underestimated correlations during summer, might correspond to cases affected by relevant snow dynamics. . . . . 79

Figure 22 Streamflow correlation decreases as the average inter-catchment distance increases. Nevertheless, the complex geomorphoclimatic heterogeneity of the study sites and the spatial variability of hydroclimatic processes result in consistent scattering (even at a short scale). Seasonal dynamics of the correlation range between river flows introduces additional variability in the relation between distance and $\rho$. Distance is computed between the center of mass of the catchments. . . . . 80

Figure 23 Up: Distribution of the exceedence frequency of the differences between normalized streamflows at couples of sites. Sites are aggregated based on their expected streamflow correlation. As model estimates of streamflow correlation decreases, inter-catchment differences in observed streamflow distributions increase (i.e larger differences between the corresponding FDCs). Down: example of representative streamflow PDFs for decreasing values of modeled correlations. From higher to lower correlations, data correspond to the following pairs of sites (USGS id) during autumn: 3451500 vs 3448000; 01048000 vs 01197500; 03136000 vs 11497500. . . . . 81

Figure 24 High values of streamflow correlation characterize pairs of catchments having similar seasonal flows across several orders of magnitudes. As streamflow correlation decreases, inter-catchment differences in average flows progressively increase. . . . . 82

Figure 25 High values of streamflow correlation characterize pairs of catchments having similar coefficient of variation $CV(q)$ across several orders of magnitudes. (i.e. similar flow regimes) . . . 83

Figure 26 High values of streamflow correlation characterize pairs of catchments having recession rates that are similar across several orders of magnitudes. As streamflow correlation decreases, inter-catchment differences in catchment response rates progressively increase. . . . . 83

Figure 27 Scatterplot between NSE and streamflow correlation for all possible pairs of seasonal streamflow timeseries obtained from the study sites. A clear relationship exists between streamflow correlation and NSE, which becomes increasingly strong as correlation increases. . . . . 84
Figure 28 Seasonal comparison of the NSE between streamflows timeseries at target-donor sites. Each study outlet is in turn assumed as ungauged (target site) and a potential reference donor site is identified by means of different criteria: i) maximizing model prediction of streamflow correlation; ii) minimizing inter-catchment distance; iii) maximizing observed streamflow correlation. When catchments are paired with the ones providing the maximum NSE, the ensuing distribution is represented by the gray shaded boxes. Model prediction of streamflow correlation systematically over performs geographic distance in the identification of reference streamgauges. 86
Figure 29 Each study site is in turn assumed as ungauged and a best reference donor streamgauge is selected. Green/red marks correspond to sites having highest NSE with the most correlated/near site. Differences in performances are evaluated in terms of the NSE between the streamflow timeseries at each pair of donor-target outlets. The size of the marks corresponds to the improve in NSE provided by the specific selection criterion compared to the other. Grey marks refer to sites where, either the two methods identify the same reference streamgauge, or none of them is able to identify a suitable donor site (NSE<0). . . . . 88
Figure 30 Comparison between the streamflow time series at a target outlet (usgs 01512500, Autumn 1957) and at two potential donor sites identified with the proximity and maximum modeled correlation criteria (usgs 01514000 and 01503000 respectively). Based on model estimates of streamflow correlation it is possible to identify a reference outlet having significantly more similar streamflow dynamics to the target outlet ($NSE_{12} = 0.62$, $NSE_{13} = -6.73$). . . . . 89

Figure 31 Distribution of the NSE (median, 45 and 55 quantiles) between the flows at pairs of catchments coupled based on maximum correlation (model prediction) and minimum geographic distance, as the average inter-catchment distance increases. It becomes more difficult to identify hydrologically similar locations when the density of a streamflow gauging network decreases (i.e. the average distance to the closest gauged site increases). However, model predictions of streamflow correlation systematically outperform spatial proximity in the identification of reference steramgauge to be associated to an arbitrary ungauged location. Differences in the performances of the two methods increase as the density of the gauging network decreases. . . . . 90

# LIST OF TABLES


Table 1 Summary of the parameters . . . . . 19
Table 2 Summary information about the study catchments . . . . . 25
Table 3 Summary of the parameters . . . . . 41
Table 4 Summary information about the study catchments . . . . . 44

# 1

# INTRODUCTION

Life relies on water. Living organisms are mainly constituted by water and in wet environments life originated and still prospers. Water as a synonym of life is a fundamental principle transcending our planet. In space exploration, scientists constantly seek for liquid water since its presence is necessary for extraterrestrial planets to harbor life. On Earth, ancient civilizations have thrived along coastlines and rivers. During the entire human history seas and rivers have provided food and drinking water, as well as a network for communication and trade. Moreover, freshwater is the main ingredient in a variety of ecological habitats, yet it is involved in many physical and environmental processes.

Gravity is the driving force that conveys precipitation – the primary input of the hydrological cycle – through hillslopes to rivers, where rain eventually feeds streamflows. While delivering streamflows to oceans, rivers shape and are shaped by the landscape and the climate of the continents they cross [*Gasparini et al., 2007; Tucker and Bras, 1998*]. In that, rivers embody the complexity and the variety of processes responsible for the continental cycle of water. The timing and the quantity of river discharge is the byproduct of meteorological and climatic dynamics (e.g. rainfall, solar radiation, wind, temperature). Streamflow dynamics at-a-station depends on the rainfall collected by the catchment (i.e. the upstream drainage area), that is neither up-taken by plants nor vaporized [*Milly, 1994; Budyko 1974*]. The sensitivity of a catchment to precipitation inputs depend, in fact, on complex and interrelated geomorphoclimatic and ecological processes. Vegetation plays a critical role in the hydrological budget as large fractions of the incoming rainfall is normally released as vapor through seasonal dynamics of plants evapotranspiration [*Porporato et al., 2004*]. Flow dynamics at-a-station additionally depend on the hydrologic response characteristic of the upstream area. Morphological and geological traits do not only control the volumes of water that can be retained by catchments. They are also critical in defining the rate at which catchments release water as a response to rain events [*Pfister et al., 2017; Chiverton et al., 2014*]. Slopes and geometry of the drainage network, as well as the hydraulic conductivity and heterogeneity of the aquifers, are critical agents to define catchment responsiveness [*Biswal and Marani, 2014; Rinaldo and Rodriguez-Iturbe, 1996*]. Impervious soils are likely to enhance the responsiveness of a catchment to precipitation inputs since they increase the fraction of rainfall that quickly drains over the land surface. However, at the

same time, poorly pervious soils are likely to retain longer the fraction of water that infiltrates. As a result, poorly pervious aquifers are prone to sustain low flows for extended periods. On the other hand, permeable soils facilitate rainfall infiltration, reducing the non-linearity and the threshold-like response of catchments [*Blöschl et al., 2013*]. The response time of a catchment is additionally controlled by the pathways that the single raindrops follow on their way to the outlet. Slow responses correspond to flow paths that drive rain water through deep, poorly pervious soil layers or bedrocks. Conversely, the most dynamic component of river flows is typically associated to shallow or superficial runoff [*Rinaldo et al., 2015*].

Streams and river systems are complex and dynamic manifestations of the hydrological cycle. While collecting and delivering freshwater across continents, they provide habitats for a wide spectrum of riparian and aquatic biota. Rivers are not only water fluxes in the hydrological budget: by transporting sediments and nutrients, they control fundamental processes that shape the morphological and ecological status of entire ecosystems. River systems are the migration network for fishes and other aquatic species and they are the pathways for the transport of nutrients, pollutants and waterborne diseases [*Bertuzzo et al., 2008*]. Moreover, their topological structure – as well as the spatiotemporal variability of discharges therein – is fundamental for the preservation of biodiversity through habitat connectivity [*Pringle, 2003*].

As many other species, humans have long benefited from flowing waters. Streams and rivers are fundamental source for freshwater, food, energy and a trade avenue. Additionally, the contribution provided by pristine environments to the well-being of humans has been acknowledged in recent years [*Berman et al., 2008; Kaplan, 1995*]. In times characterized by increasing psychological stress caused by a chaotic and fast evolving society, the importance of natural environments – especially within densely urbanized areas – is becoming evident. Consequently, ecosystem services related to the good ecological status of rivers is receiving increasing interest (e.g. `http://www.alpine-space.eu/projects/hymocares/en/home`). Nonetheless, major challenges are ahead, as the technical progress of the modern age have severely compromised the integrity of many habitats.

Since more than a century, the potential energy of waters is widely harvested causing major threats to the continuity of riverine ecosystems. Hydropower, despite providing one of the most sustainable high-yield forms of energy, heavily affects the natural dynamics of streamflows [*Poff et al., 1997*]. Moreover, impoundments represent an obstacle for riverine habitat connectivity and for the continuity of sediment transport along stream networks. The hydrological cycle is also sensitive to human-induced modifications of the water balance (i.e. the volume of water released by a catchment in a given period of

time) and of landscape response rates (i.e. how quickly runoff reaches a catchment outlet). In particular, land use shifts, combined with climate change and river channelization are responsible for the alteration of the feedback loop between hydrological forcing and streamflow dynamics [*DeFries and Eshleman, 2004*]. Furthermore, enhanced variability of streamflow conditions caused by an excessive anthropogenic pressure on the environment ultimately increases the risk of extreme events like floods and droughts in many regions of the world.

Across all Europe, anthropogenic-induced alterations of flow and morphological dynamics have dramatically impaired the potentials of many rivers to provide ecological habitats and to sustain biodiversity. Nowadays, many river systems are reduced to mere channels devoted to convey water as effectively and fast as possible. In most developed countries, heavy engineered rivers are the heritage of decades featured by poor environmental sensitivity. In the past, land reclamation – a priority over the natural integrity of the landscape – urged straight channels with compact sections in order to secure the surrounding areas from the risk of flooding. Fortunately, a new tendency has started establishing in recent years. Rivers are acknowledged to be something more than mere pipes, but rather complex systems providing a wide spectrum of ecosystem services. Evidences on the benefits provided by pristine river systems to water quality, habitat diversity and human psychological health have recently catalyzed the interest over river restoration practices [*Wohl et al., 2015*]. Research is conducted in many national and regional institutions and improvements in restoration practices have benefited from the consistent investments that many countries are devoting to the study of the physical and biogeochemical functioning of river systems (e.g.:http://www.bayceer.uni-bayreuth.de/hypotrain/).

The European Union's Water Framework Directive (WFD) is a significant example of how the rising awareness and concern upon the ecological status of water bodies has triggered new environmental policies (http://ec.europa.eu/environment/water/water-framework/index_en.html). Together with other communitarian directives concerning waters (e.g. urban wastewater directive, nitrate directive, drinking water directive), the WFD represents a joint effort to ameliorate the biochemical and morphological status of many degraded European water bodies. The ambitious goal of the WFD is to unify country-specific regulations towards higher standards while raising the awareness of the society on the health conditions of aquatic environments. The WFD represents the first attempt to comprehensively evaluate the ecological status of river networks by including river morphology among the classical chemical and biological indicators of water quality. Consequently, rivers in good ecological status combine the lack of biochemical contamination with naturally occurring morphologi-

cal heterogeneities in their streambed, riparian corridors and floodplains. However, in addition to potential practical and financial constraints, the implementation of new strategies to cope with human-impacted streams raises conceptual challenges. After centuries of anthropogenic pressures on the environment, defining "natural" ecological and morphological conditions can be problematic [*Carlisle et al., 2009*]. Therefore, hydrographic network urge to be studied within a holistic approach that accounts for spatial variability of discharge and sediment dynamics. Moreover, catchment-scale controls of river morphology (e.g. flow regimes, land use, sediment transport) need to be jointly included in the definition of the ecological potential of aquatic ecosystems.

Risk prevention strategies as well as ecosystem services provided by habitat integrity and recreational opportunities strictly depend on the temporal variability of flow conditions along river systems. Additionally, accounting for the dynamics of river flows is crucial for the optimal management of water resources [*Razurel et al., 2018; Niayifar and Perona, 2017*]. Accurate estimates of the streamflow volumes available in specific periods of the year is necessary to design reservoirs devoted to flood control or hydroelectric production, as well as in the design of run-of-river power plants [*Lazzaro et al., 2013; Basso and Botter, 2012*]. The probabilistic nature of hydrological forcing, and the ensuing variability of river discharge, is particularly visible during extreme flows in erratic or flashy rivers [*Botter et al., 2013*]. Understanding streamflow dynamics and the underpinning hydrological processes is thus crucial in managing the risk related to floods and droughts and for the interpretation of morphoclimatic legacies on the geochemical functioning of river network.

Nonetheless, the complexity of large scale climatic and ecological patterns, combined with the uncertainties related to small-scale geological and morphological heterogeneity, often hinder the functional link between geomorphoclimatic forcing and the hydrological response of river basins. As a consequence, conceptual and observational efforts are demanded to disentangle and characterize the drivers of flow regimes along river networks.

The probability density function (PDF) of streamflows at one river section is an effective tool to capture relevant statistical properties of flow regimes. Once streamflow measurements are available, empirical PDFs can be directly evaluated based on the frequency a discharge value is recorded during a given period. The period of interest can include a single year (or a single season) or an ensemble of years (or seasons), depending on the interest focus. For example, when the population of yearly PDFs is considered, the inter-annual variability of flow statistics can be investigated [*Dralle et al., 2017*].

Unfortunately, direct streamflow measurements are usually available at a limited set of unevenly distributed locations along river

networks. In addition, the recorded timeseries can be too short to perform statistically-sound inference. The lack of an adequate monitoring of hydrological variables is especially critical in undeveloped or in developing countries. Therein, inadequate knowledge of river flow conditions can strongly interfere with the economical and industrial development, and can severely impact the quality and safety of human activities. Moreover, the weak resilience to unusual flow conditions can further expose these settings to physical and sanitary hazards connected to extreme flow events.

To cope with the widespread lack of direct flow measurement, methods have been developed to estimate flow characteristics in ungauged locations [*Blöschl et al., 2013*]. Statistical techniques aim to extrapolate flow statistics from gauged to ungauged sites by exploiting empirical correlations between hydrological variates [*Kiang et al., 2013; Merz and Blöschl, 2014*]. Statistical approaches can be grouped into three classes: regression models, classification-based approaches and geostatistical techniques. In the first case, streamflow statistics are estimated at ungauged sites based on regression analysis. Statistical relationships are established between available physiographic attribute and streamflow statistics in a set of catchments where streamflow is measured [*Pugliese et al., 2016*]. The relationships are then used to estimate unknown statistics at sites where physiographic attributes are available but flow data are lacking.

The second class of statistical methods adopts clustering or other classification techniques (e.g. principal component analysis) to identify catchments sharing similar geomorphoclimatic traits [*Wagener et al., 2007; He et al, 2011*]. The rationale being that the hydrological response of physiographically similar catchments is expected to be similar. Consequently, hydrological information can be exported from gauged to ungauged site, provided their membership to the same class.

Geostatistics is the oldest approach to evaluate streamflow signatures in absence of discharge data [*Vandewiele et al., 1991; Blöschl, 2005; Mohamoud, 2010*]. Geostatistical methods have long been employed in many earth sciences to evaluate spatial fields of geophysical variables [*De Marsily, 1986*]. In hydrology, these class of techniques aim to extrapolate across space the information obtained from pointwise streamflow measurements. Geostatistical methods assume that the physical processes underlying the hydrological response of the landscape are smoothly variable across space. As a consequence, neighboring catchments are expected to share similar hydrological characteristics. In the simplest case, streamflow signatures at an ungauged location are obtained by exporting the corresponding signature measured at the closest steam gauge. Different schemes can be used to weight the contributions of multiple neighboring streamgauges (e.g. inverse distance). Kriging is an unbiased refinement of

these techniques. While weighting the contributions of the different donor streamgauges, kriging accounts for the spatial autocorrelation structure of the signature being estimated. Nevertheless, in assuming isotropy, popular kriging schemes might overlook important legacies on flow regimes deriving from the morphological arrangement of river catchments. As a result, specific algorithms have been developed that account for the topological structure of the hydrographic networks to provide a more realistic interpolation of flow-related statistics [*Skøjen et al., 2006*].

Statistical methods are powerful tools to describe the spatial variability of river flow conditions along and across river systems [*Pugliese et al., 2016; Castellarin, 2004*]. Together with other data driven approaches (such as neural networks, deep learning and artificial intelligence [*Rao, 2000*]), these techniques are promising to unravel relationships between hydrological variables in a context of increasing data availability (e.g.: remotely sensed data). Moreover, statistical approaches can also be used to estimate and export entire sets of model parameters [*Beck et al., 2016*]. In this prospect streamflow dynamics at an ungauged location can be computed using rainfall-runoff models combined with regionalized model parameters that have been previously calibrated on available gauged sites.

Nevertheless, when the lack of runoff monitoring precludes the use of such data-intense approaches, more parsimonious methods are necessary. These methods compensate for the reduced information deriving from scarce observational data with a mechanistic description of the processes underlying flow dynamics [*Doulatyari, 2016*]. Physically based approaches are less flexible than statistical methods as they rely on a set of modelling assumptions. However, compared to statistical methods, physically based models generally provide more robust predictions and allow a direct interpretation of catchment functioning by linking drivers and responses through causal laws.

Stochastic physically-based approaches attempt to link flow variability to the probabilistic nature of catchment-scale hydroclimatic and geomorphological forcing. They can provide analytical expressions for streamflow PDFs as a function of lumped parameters describing, for example, the frequency and intensity of runoff-generating rainfall events [*Botter et al., 2007a; Doulatyari et al., 2015*]. Despite the limited data required and the strong modeling assumptions, these models are effective to capture the steady-state variability of flow regimes.

Spatially distributed numerical algorithms would, in principle, be more accurate in resolving the spatial heterogeneity of aquifer storage and transport processes. Numerical methods couple and solve at fine spatial resolution the fundamental water balance and flux laws typical of flow processes in porous media. Despite the large data requirement, also spatially distributed models generally offer a simpli-

fied representation of fundamental processes and variables, such as recharge patterns and hydraulic conductivity. The limited knowledge of the system and the complexity of processes that can hardly be reduced to a set of partial differential equations, challenges the "spatially explicit" and "physically realistic" nature of such models. For example, using well-defined flow domains with deterministic, homogeneous and histotrophic hydraulic conductivities or boundary conditions might, in some cases, be unrealistic [*Blöschl et al., 2016*]. Moreover, the predictive uncertainties deriving from commonly adopted assumptions are typically difficult to assess within spatially distributed numeric frameworks. Uncertainty analysis based on statistically significant ensembles of model realization would in fact be time consuming and computationally demanding, even for steady-state simulations. As a result, also sophisticated spatially-distributed numerical models require complex processes to be conceptualized. In particular, the explicit coupling of physical processes across different flow domains is especially critical and is handled by a limited number of models. Hence, processes such as the interaction between surface water and groundwater through the streambed (e.g. groundwater recharge or hyporheic exchange), are often quantified via simplified steady-state approaches [*Boano et al., 2014*].

Establishing computational intensive numerical model can be worthwhile for addressing specific research questions or for relevant engineering purposes in well-defined domains (e.g. to identify the flow paths intercepted by a drinking water supply facility). However, more parsimonious methods may represent valuable alternatives when the focus is on large scales, particularly in poorly gauged areas.

This thesis presents a new approach to characterize spatial patterns of flow regimes and to predict streamflow signatures in ungauged locations. Through the description of the main catchment-scale physical processes responsible for streamflow dynamics, a quantitative analysis is performed on the hydrological controls of flow regimes. In particular, the study focuses on the spatial correlation between synchronous daily streamflow timeseries at two arbitrary locations along a river network. Results show, indeed, that streamflow correlation is a synthetic yet effective indicator of hydrological similarity. Therefore, a mechanistic quantification of flow correlation can be used to evaluate the spatial variability in the hydrological response of river basins as a function of the underlying hydrological drivers.

Streamflow cross-correlation is predicted by analytically coupling the flow dynamics at two catchment outlets. Flow dynamics at the two sites are stochastically modeled as a function of the frequency and intensity of the daily effective rainfall events (i.e. runoff-generating rainfall) in the corresponding contributing areas. Recession rates of the hydrographs are additionally considered to account for the catchments response times. Frequency and intensity of effective rainfall

events characterize the hydrologic forcing on streamflow dynamics and quantify the ensuing water balance, whereas recession rates define how fast the effective components of rainfall inputs are released by the catchments.

The timing of the effective rainfall events in the two catchment is crucial for the coupling of flow dynamics and, consequently, for the quantification of streamflow spatial correlation. For this purpose, the modeling framework distinguishes between two classes of effective rainfall events: joint and disjoint. Joint effective rainfalls trigger simultaneous streamflow increments at the outlet of both catchments. Conversely, disjoint effective rainfalls produce runoff at just one catchment at the time. This distinction is paramount since the similarity in the hydrological response of two catchments is critically affected by the relative frequency and intensity of joint events.

The parsimonious modeling framework suffices to provide robust estimates of the spatial correlation between seasonal streamflow timeseries across the wide range of geomorphoclimatic conditions offered by the contiguous USA. The framework is tested on a large scale to predict correlation and other hydrological signature across 413 American catchments. A version of the model is additionally developed to identify pairs of river outlets characterize by similar hydrological response without requiring any calibration on available streamflow data. This version of the model – which only requires rainfall data – proves appealing for predicting streamflow signatures and hydrological similarity in large, poorly gauged areas.

In summary, this thesis provides a methodology to study spatiotemporal traits of flow regimes through the characterization of streamflow mutual correlation. By quantifying spatial patterns of flow dynamics, the framework offers a mechanistic basis to assess the effect of geomorphoclimatic controls on river flows and to assess the impact of different hydrological drivers on the spatial patterns of flow regimes. The model can help interpreting the hydrological legacies on the biogeochemical functioning of river systems and represents a tool for optimizing the spatial distribution of streamgauges and hydroelectric facilities along hydrographic networks.

## 1.1 AIMS AND OBJECTIVES

The scope of this work was to develop a new framework to quantify hydroclimatic and ecomorphological legacies on the spatial patterns of streamflow regimes. The goal is pursued through the development of a catchment-scale analytical model. The model accounts for the timing and the intensity of effective rainfall events – as well as for the catchment response rates – to quantify the correlation between daily streamflows at two arbitrary locations. Since streamflow correlation is found to be a powerful indicator of the similarity between the hydro-

logical responses of two catchments, the model is a tool to interpret spatial patterns of flow dynamics along river networks. Moreover, model predictions of streamflow correlation in absence of discharge data are appealing for practical and scientific purposes such as flow predictions in ungauged catchments. More specifically, the research objectives of this work are:

- through a probabilistic approach, to develop a physically-based analytical model that quantifies the Pearson correlation between daily streamflows at arbitrary pairs of outlets as a function of a limited set of catchment-scale hydrological processes;
- to analyze the sensitivity of flow correlation to inter-catchment differences in the processes controlling streamflow dynamics at the two outlets;
- via a large-scale application, to assess the performances of the model in predicting streamflow correlation across strong geomorphoclimatic gradients;
- to implement a version of the model that can be used to predict streamflow correlation between arbitrary locations along river networks in absence of discharge data without requiring calibration over existing flow records;
- to assess the information embedded in the streamflow correlation as an index of similarity between the hydrological responses of two catchments.
- to evaluate the performances of correlation as a regionalization criterion for streamflow signatures. In particular, to assess how correlation compares to inter-catchment distance in the estimation of streamflow timeseries at ungauged locations.

## 1.2 THESIS STRUCTURE

The results of the outlined research are presented as follows:

CHAPTER 2 develops the stochastic model that will be used to characterize and predict the spatial correlation of daily flows. After an encouraging proof of concept, the analytical model is analyzed. The theoretical sensitivity of streamflow correlation to inter-catchment differences in the processes considered in the analytical description of flow dynamics is evaluated.

CHAPTER 3 presents a regional-scale application of the method. An additional version of the model is implemented that doesn't require calibration over existing streamflow records. Model parameters are evaluated based on observational data to assess

how the heterogeneity of the physical processes responsible for streamflow dynamics ultimately affect streamflow correlation in real cases.

CHAPTER 4 includes a comprehensive evaluation of streamflow correlation as a metric of hydrological similarity. The relationship between streamflow correlation and the other signatures that define the hydrological response of river basins is assessed. Interseasonal dynamics in the spatial patterns of flow regimes are highlighted as they are critical for the selection of hydrologically similar outlets. Finally, model predictions of streamflow correlation are used in absence of discharge data to identify location with similar flow dynamics. The regionalization performances based on streamflow correlation and on spatial proximity are compared.

CHAPTER 5 concludes the study summarizing pivotal findings, limitations of the model, and future challenges in the prediction of spatial patterns of flow regimes. Research perspectives are delineated together with the broad range of practical and scientific opportunities provided by the methods presented in this study.

# 2

# CHARACTERIZING THE SPATIAL CORRELATION OF DAILY STREAMFLOWS

Andrea Betterle[1,2,3], Mario Schirmer[1,2], Gianluca Botter[3]
Water Resources Research, 2017

ABSTRACT

In this study we propose an analytical framework to estimate the spatial correlation of daily flows in two arbitrary locations within a given hydrologic district or river basin. The method builds on the description of the coupled streamflow dynamics at the outlet of two catchments, which are represented as correlated shot noises forced by Poisson rainfall. Novel analytical expressions for the spatial correlation of discharge are derived using a limited number of parameters that encapsulate effective precipitation regime and catchment drainage rates. The method is suited to describe how heterogeneity of climate and landscape features impact the spatial and temporal variability of flow regimes along river systems. The analysis suggests that frequency and intensity of synchronous effective rainfall events in the relevant contributing catchments are the main driver of the spatial correlation of daily discharge, unless the drainage rates of the two catchments differ by almost one order of magnitude. The method also portrays how the topological arrangement of the two outlets along the river network influences the underlying streamflow correlation, and shows how nested catchments tend to maximize the spatial correlation of flow regimes. To demonstrate the potential of the tool, the model is tested on a set of sixteen catchments belonging to a *120,000* $km^2$ region of the United States. The application evidences satisfactory performance ($RMSE < 0.1$) in reproducing the spatial correlation of daily streamflows among the study sites. The approach provides a clue for the characterization of water availability in space at seasonal time scale, with implications for water resources assessment, risk prevention and ecological studies.

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

## 2.1 INTRODUCTION

A proper characterization of streamflow patterns in space and time represents a significant scientific challenge with a wide range of implications for human water uses and ecosystem services conservation [*Postel and Richter, 2003; Ziv et al., 2012; Hurford and Harou, 2014*]. Despite this, relatively few rivers are adequately monitored [*Blöschl et al. 2013; Kiang et al., 2013*] and improving the density of existing gauging networks is often challenged by technical and economical limitations, that include the availability of financial resources and the accessibility of stream reaches. Therefore, in most practical settings, observational data about spatial and temporal patterns of flow regimes may be inadequate for water resources management and for the prediction of risk associated to floods and droughts.

To cope with the absence of dense discharge gauging networks, many different approaches have been proposed in the literature to predict streamflow availability in sparsely gauged or ungauged catchments (see e.g. *Blöschl et al. [2013]* and references therein). Probabilistic models are typically concerned with the frequency distribution of discharge (PDF) or with the corresponding flow duration curve (FDC), thus disregarding temporal dynamics of flows. Although certain probabilistic approaches (e.g. *Botter et al. [2008a], Müller et al. [2014]*) entail a physically-based mechanistic formulation that facilitates hydrologic prediction in sparsely gauged catchments and under changing climatic conditions [*Doulatyari et al., 2015; Müller et al., 2015*], cross-correlations among multiple outlets have never been studied in that framework. On the contrary, spatial patterns of streamflow regimes are usually accounted for in regionalization methods [*Castellarin et al., 2004; Castellarin et al., 2007; Cheng et al., 2012*], which attempt to establish explicit connections among gauged and ungauged sites based on the concept of hydrologic similarity. In particular, geostatistical methods have proven effective in describing spatial patterns of river flows. For example, Top-Kriging [*Skøjen et al., 2006; Castiglioni, 2011; Laaha, 2014*] is a topological approach for the interpolation of flow-related variables along river networks that explicitly considers the spatial arrangement of catchments. The *Map Correlation Method* [*Archfield and Vogel, 2010*] is another geostatistical approach devoted to the selection of the gauged section that is most correlated with the target ungauged site. Streamflow correlation, in fact, has been proven to be a better indicator than spatial proximity for transferring flow attributes from gauged to ungauged sites.

Spatially explicit numerical rainfall-runoff models [*Zehe and Blöschl, 2004; Rigon et al., 2006; Kollet and Maxwell, 2008; Costa-Cabral et al., 2008; Schaefli et al., 2014*] have the potential to provide a detailed characterization of the spatial and temporal patterns of runoff hydrographs by exploiting information about catchment characteristics

and relevant climatic features. However, numerical models require intensive calibration and the application to ungauged outlets remains challenging [*Blöschl et al., 2006; Castiglioni et al., 2010*].

Notwithstanding the progresses made in the characterization of flow regimes, the understanding of spatial connections in streamflows is an area where more research is needed [*Sivakumar and Woldemeskel, 2014*]. In this context, exploring the spatial and temporal structure of the streamflows correlation based on simple climate and hydrologic attributes represents an attractive prospect. The correlation of the streamflows is a synthetic statistical descriptor of similarity between synchronous discharge dynamics at arbitrary pairs of outlets. Therefore it encapsulates the complex effect of heterogeneous climate, geology, and land cover on flow regimes [*Kiang et al., 2013*]. Highly correlated outlets are likely to display similar discharge dynamics (Figure 1), thereby enabling for a more efficient extrapolation of available streamflow records. For example, a proper characterization of the spatial correlation of streamflows supports the prediction of flow statistics in ungauged sections [*Archfield and Vogel, 2010; Messinger and Paybins, 2014*] and facilitate the identification of optimal configurations for water infrastructures along river networks. A physically based characterization of discharge correlation could also contribute to improve the reliability of the estimate of long-term streamflow statistics in sites where only short-term records are available (e.g. via record augmentation techniques [*Hirsch, 1982; Vogel and Stedinger, 1985*]).

A better understanding of spatial connections in streamflow dynamics could be used to more efficiently expand existing hydrometric gauging networks [*Kiang et al., 2013*]. In fact, predicted streamflow correlations could be exploited to identify locations that are poorly correlated with existing gauging stations and, thus, are best suited to be equipped with new stream gauges. This would eventually support the optimization of geostatistical techniques designed for the spatial interpolation of observed flow attributes along river networks.

Following a stochastic framework, this work aims to provide a physically-based analytical characterization of the seasonal correlation coefficient between daily streamflows at the outlet of two arbitrary catchments (nested or disjointed). The steady-stated, zero lag streamflow correlation is expressed as a function of lumped parameters that embody seasonal climatic and landscape features (rainfall, soil, vegetation, recession rate) in the contributing areas.

The remainder of this manuscript is organized as follows: Section 2.2 describes the stochastic approach used to derive a general analytical expression for the spatial correlation of daily streamflow. In order to make the general solution operative, some assumptions have to be made about the probability distribution of the effective rainfall depths. For this purpose, a set of alternatives is presented and dis-

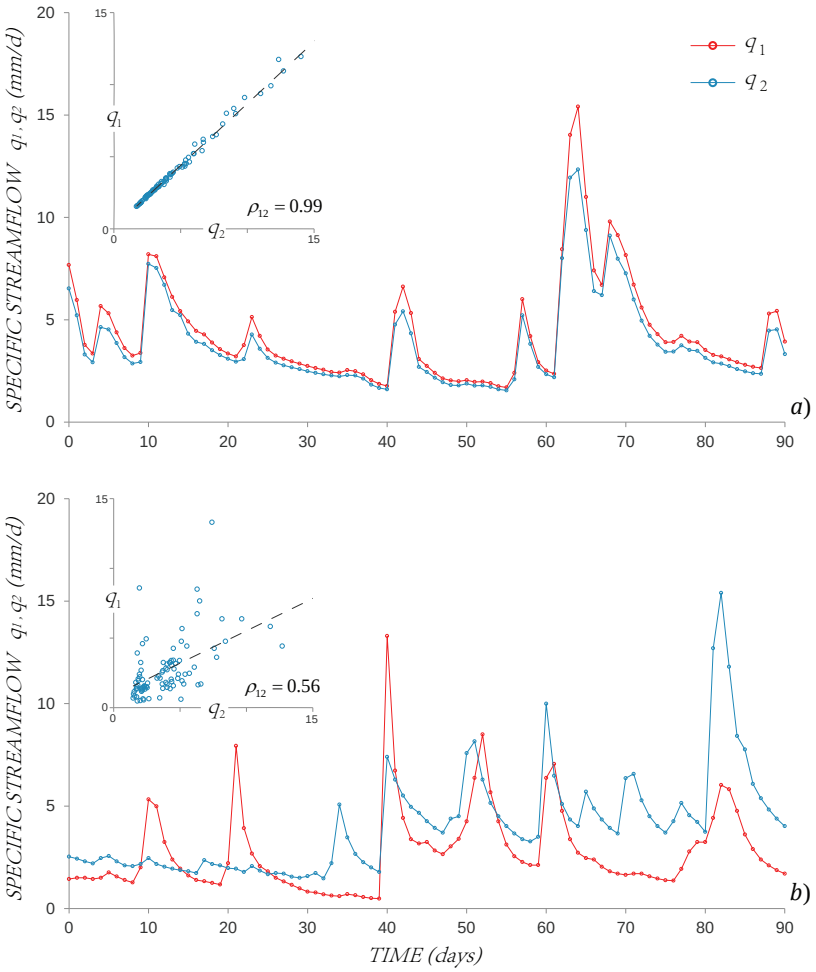


Figure 1: Comparison between the streamflow timeseries at two pairs of catchment outlets in the Eastern United States. Figure 1a: Little Pigeon River at Sevierville (TN) ($q_1$) and Oconaluftee River at Birdtown (NC) ($q_2$), during autumn; Figure 1b: Valley River at Tomotla (NC) ($q_1$) and North Fork Holston River near Gate City (VA) ($q_2$), during spring. The insets show the scatter plots of $q_1$ and $q_2$.

cussed in Section 2.3, jointly with the derivation of the corresponding analytical expressions of the streamflow spatial correlation. The performances of the model are assessed in Section 2.4, where the model is applied to a set of case studies. A detailed sensitivity analysis is presented in Section 2.5, while the impact of heterogeneity of rainfall, soil and vegetation properties on the streamflow correlation is further discussed in Section 2.6. A set of conclusions closes the paper.

## 2.2 ANALYTICAL CHARACTERIZATION OF STREAMFLOW CORRELATION

In order to develop an analytical expression for the spatial correlation of streamflows at the seasonal timescale, we first specify the joint dynamics of daily discharge at two arbitrary catchment outlets during that season. In rivers that are not affected by relevant water storage (e.g. lakes, reservoirs, snowpacks) abrupt increases of discharge (streamflow jumps) result from the random occurrence of flow-producing (i.e. effective) rainfall events [*Claps et al., 2005; Botter et al., 2007a; Müller et al., 2014; Andres-Domenech et al., 2015; Doulatyari et al., 2015*]. Each streamflow jump is then followed by a recession resulting from the drainage of the contributing catchment (Figure 1).

The effective rainfall within each catchment is described as a marked Poisson process $\xi_{it}$ with frequency $\lambda_{it}$ (where the subscript $i \in \{1, 2\}$ identifies the relevant contributing catchment, and $t$ refers to the "total" series of effective rainfall events). Poisson processes in time describe the occurrence of independent events with exponentially distributed interarrivals [*Kingman, 1992*]. In this case, the marks of the process are represented by the effective rainfalls depths in catchment $i$, $h_i$, that are assumed to be random variables characterized by the probability density function (PDF) $b_{it}$. The frequency of rainfall events is higher than the frequency of runoff producing events because of the ability of the root zone to buffer incoming rainfall during wetting-drying cycles. The buffering capacity of the soil crucially depends on the soil water storage capacity, land cover and climate [*Milly, 1994; Porporato et al., 2004; Botter et al., 2007a; Botter et al., 2013; Thompson et al., 2011*]. The parameters $\lambda_{it}$ hence summarize climate and landscape attributes in the contributing catchments and encapsulate the major nonlinearities in rainfall-runoff transformation.

We also assume that the daily specific discharge (per unit of catchment area) at the considered outlets ($q_1$ and $q_2$) can be expressed as the convolution between the effective rainfall ($\xi_{it}$) and an exponential unit hydrograph:

$$\begin{cases} q_1(t) = \int_0^t \xi_{1t}(t-\tau) k_1 \, e^{-k_1 \tau} \, d\tau \\ \\ q_2(t) = \int_0^t \xi_{2t}(t-\tau) k_2 \, e^{-k_2 \tau} \, d\tau \end{cases} \tag{1}$$

where $k_i$ identifies the streamflow recession rate in the catchment $i$ (i.e. the inverse of the mean response time of the hydrograph). In spite of their simplicity, exponential hydrographs have been successfully employed to characterize streamflow statistics in a wide range of geographical and climatic settings [*Botter et al., 2007c, 2013; Pumo*

*et al., 2013; Müller and Thompson,2015*]. The coupling between the dynamics of $q_1$ and $q_2$ emerges from the synchronous effective rainfall events in the two catchments, resulting in the terms $\xi_{1t}$ and $\xi_{2t}$ in equation (1) being correlated.

To investigate the correlation between $q_1$ and $q_2$ it is necessary to introduce some additional parameters that describe the climate forcing within the two contributing catchments. In particular, $\lambda_1$ ($\lambda_2$) identifies the frequency of effective rainfall events that generate streamflow only in the catchment 1 (2) (disjoint events). Conversely $\lambda_{12}$ identifies the frequency of rainfall events determining a simultaneous streamflow increment in both the outlets (joint events). Because of the low autocorrelation of daily rainfall [*Zorzetto et al., 2016*], joint and disjoint events are assumed as independent Poisson processes. The total frequency of runoff producing events at each catchment outlet is therefore $\lambda_{it} = \lambda_i + \lambda_{12}$, ($i = 1, 2$). Additionally, let us define the PDF of the intensity of disjoint effective rainfall events in the catchment "$i$" as $b_i(h_i)$. Likewise, $b_{12}(h_1, h_2)$ is the bivariate PDF characterizing the depths of effective rainfall when streamflows are simultaneously generated in both catchments (see Table 1 for a detailed list of the notation used in this paper). In summary, the total effective rainfall in equation (1) can be expressed as the sum of two stochastic noises, as detailed below (the first argument of each term indicates the average frequency of the effective rainfall events, while the second refers to the PDF of their intensities):

$$\begin{cases} \xi_{1t}\left[\lambda_{1t}; b_{1t}\right] = \xi_1^{12}\left[\lambda_{12}; b_1^{12}\right] + \xi_1\left[\lambda_1; b_1\right] \\ \\ \xi_{2t}\left[\lambda_{2t}; b_{2t}\right] = \xi_2^{12}\left[\lambda_{12}; b_2^{12}\right] + \xi_2\left[\lambda_2; b_2\right] \end{cases} \tag{2}$$

The first terms on the right hand side of equations (2), $\xi_i^{12}$, are related to the streamflow jumps produced by joint effective rainfall events, while the second terms ($\xi_i$) account for the streamflow jumps occurring only at one catchment outlet (disjoint events). In equation (2), the marginal PDF of effective rainfall depths in the catchment $i$ produced by joint events, $b_i^{12}$, can be calculated from the corresponding joint PDF ($b_{12}$) as: $b_1^{12} = \int_0^\infty b_{12}(h_1, h_2)\, dh_2$; and $b_2^{12} = \int_0^\infty b_{12}(h_1, h_2)\, dh_1$. Figure 2 graphically represents the decomposition of the total effective rainfall into joint and disjoint events and the corresponding streamflow dynamics produced at the outlet of each catchment, according to equations (1) and (2). Note that, in the proposed formulation, we neglect any kinematic delay between the streamflow jumps observed at the considered catchment outlets in response to joint events. The above simplification is applicable whenever the timescale of flood waves propagation along the river network is shorter than the temporal resolution adopted in this study (1 day).

As a rule of thumb, in case of an ideal circular catchment with a maximum channel length equal to its diameter, if we assume a wave celerity of *3 m/s* and accept a maximum delay between the hydrograph peaks of *10* hours, the assumption holds for catchment areas $A \lesssim 10^4$ km$^2$.

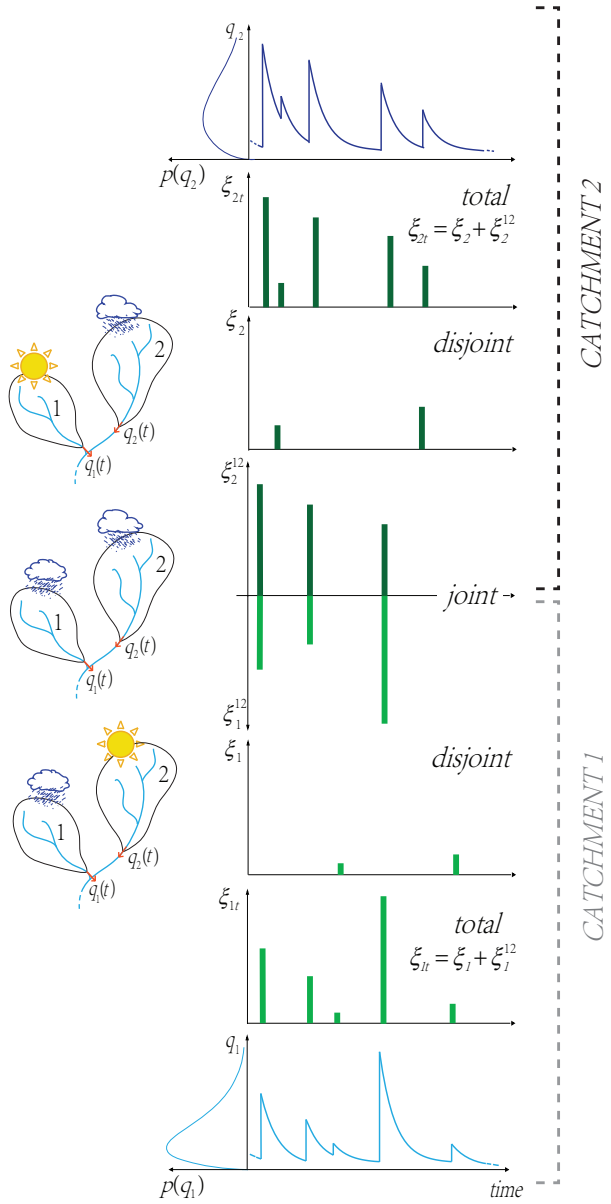


Figure 2: When a pair of catchments are simultaneously considered, the overall timeseries of effective rainfall ($\xi_{it}$) within each catchment can be decoupled into two sub-timeseries. The first includes the joint effective rainfall events (when streamflow jumps are observed at the outlet of both catchments simultaneously) $\xi_i^{12}$, whereas the other includes the effective rainfall events experienced only by one of the two catchments, $\xi_i$.

The master equation for the joint probability density function of the streamflows $q_1$ and $q_2$ associated to equation (1) can be written as (e.g.: *Gardiner [1983], Isham et al. [2005], Botter et al. [2008]*):

$$\begin{aligned}\frac{\partial p(q_1,q_2,t)}{\partial t} &= \frac{\partial\left[k_1 q_1 p(q_1,q_2,t)\right]}{\partial q_1}+\frac{\partial\left[k_2 q_2 p(q_1,q_2,t)\right]}{\partial q_2}+\\ &+ \lambda_1\int_0^{\frac{q_1}{k_1}} b_1(h_1)\,p(q_1-k_1h_1,q_2,t)\,dh_1+\\ &+ \lambda_2\int_0^{\frac{q_2}{k_2}} b_2(h_2)\,p(q_1,q_2-k_2h_2,t)\,dh_2+\\ &+ \lambda_{12}\int_0^{\frac{q_1}{k_1}}\int_0^{\frac{q_2}{k_2}} b_{12}(h_1,h_2)\,p(q_1-k_1h_1,q_2-k_2h_2,t)dh_1dh_2-\\ &- (\lambda_1+\lambda_2+\lambda_{12})\,p(q_1,q_2,t)\end{aligned} \tag{3}$$

Equation (3) states that the temporal variation of the joint probability of the streamflows $q_1$ and $q_2$ can be expressed as the sum of six independent terms: the gain/loss of probability due to the deterministic decay of discharge at the outlets 1 and 2 during recessions; the increment of probability due to disjoint events in catchment 1 and 2; the increment of probability due to joint events; the loss of probability due to streamflow producing events (regardless of their intensity and nature).

The steady-state solution for equation (3) in terms of the moment generating function (MGF) of the joint PDF of $q_1$ and $q_2$ ($\hat{\hat{p}}(s_1,s_2)$, being $s_1$ and $s_2$ the Laplace variables associated to $q_1$ and $q_2$) reads (e.g. *Van Kampen [1992], Isham et al. [2005], Botter et al. [2007b]*):

$$\begin{aligned}\hat{\hat{p}}(s_1,s_2) &= \exp\Bigg\{-\lambda_1\int_0^{\infty}\left[1-\hat{B}_1(k_1s_1e^{-k_1t})\right]\,dt-\\ &- \lambda_2\int_0^{\infty}\left[1-\hat{B}_2(k_2s_2e^{-k_2t})\right]\,dt-\\ &- \lambda_{12}\int_0^{\infty}\left[1-\hat{\hat{B}}_{12}(k_1s_1e^{-k_1t},k_2s_2e^{-k_2t})\right]\,dt\Bigg\}\end{aligned} \tag{4}$$

where $\hat{B}_1$, $\hat{B}_2$ and $\hat{\hat{B}}_{12}$ are the Laplace transforms of $b_1$, $b_2$ and $b_{12}$, respectively.

The correlation between $q_1$ and $q_2$ ($\rho_{12}$) can be derived from the MGF of equation (4) as detailed in the Appendix 2.8.1. The analytical expression of $\rho_{12}$ reads:

$$\rho_{12} = \frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}} \frac{\int_0^\infty \frac{\partial^2 \hat{B}_{12}(k_1 s_1 e^{-k_1 t}, k_2 s_2 e^{-k_2 t})}{\partial s_1 \partial s_2} dt \Big|_{\substack{s_1=0 \\ s_2=0}}}{\sqrt{\int_0^\infty \frac{\partial^2 \hat{B}_{1t}(k_1 s_1 e^{-k_1 t})}{\partial {s_1}^2} dt \int_0^\infty \frac{\partial^2 \hat{B}_{2t}(k_2 s_2 e^{-k_2 t})}{\partial {s_2}^2} dt \Big|_{\substack{s_1=0 \\ s_2=0}}}} \tag{5}$$

The analytical expression of the seasonal streamflow correlation (equation (5)) is given by the product between two ratios. The first is related to the frequency of effective rainfall events and it can be interpreted as a normalized frequency of joint events. The second accounts for the intensities of the events as well as for key properties of the hydrologic response in the two contributing catchments (i.e. the rate at which the two basins process the excess of rainfall). It is worth to note that equation (5) holds regardless of the specific distribution chosen for the effective rainfall depths, with the only assumption being that effective rainfall events are homogeneous Poisson processes.

Table 1: Summary of the parameters

| Parameter | Description |
|---|---|
| $\lambda_{it}$ | Average frequency of all effective rainfall events in catchment $i$ |
| $\lambda_i$ | Average frequency of disjoint effective rainfall events in catchment $i$ |
| $\lambda_{12}$ | Average frequency of joint effective rainfall events in the two catchments |
| $\lambda_m$ | Minimum between $\lambda_{1t}$ and $\lambda_{2t}$ |
| $\alpha_{it}$ | Average depth of all effective rainfall in catchment $i$ |
| $\alpha_i$ | Average depth of disjoint effective rainfall in catchment $i$ |
| $\alpha_i^{12}$ | Average depth of joint effective rainfalls in catchment $i$ |
| $k_i$ | Streamflow decay rate during recessions in catchment $i$ |
| $r_\alpha$ | Correlation between the joint effective rainfall depths in the two catchments |
| $a$ | Slope of the linear relation between the joint effective rainfall depths |
| $b_{it}(h_i)$ | PDF of all effective rainfall depths in catchment $i$ |
| $b_i(h_i)$ | PDF of disjoint effective rainfall depths in catchment $i$ |
| $b_{12}(h_1, h_2)$ | Bivariate PDF of joint effective rainfall depths in the two catchments |
| $b_i^{12}(h_i)$ | Marginal PDF of joint effective rainfall depths in catchment $i$ |
| $\hat{B}_{it}(s_i)$ | Laplace transform of $b_{it}$ |
| $\hat{B}_{12}(s_1, s_2)$ | Laplace transform of $b_{12}$ |

i=1,2 identifies the two catchments;
joint: affects at the same time both catchments;
disjoint: affects only one catchment at a time.

## 2.3 EXPONENTIALLY DISTRIBUTED EFFECTIVE RAINFALL DEPTHS

To make equation (5) operative, the analytical expressions for the PDFs characterizing the effective rainfall depths ($b_{1t}$,$b_{2t}$ and $b_{12}$) need to be defined. In this section we specify equation (5) for exponentially distributed effective rainfall depths. Exponentially dis-

tributed rainfall depths have been frequently employed in analytical studies focused on the impact of stochastic rainfall on the water cycle [*Rodriguez-Iturbe et al., 1999; Laio et al., 2001; Botter et al., 2007a, 2008; Verma et al., 2011; Basso et al., 2015, 2016*].

The bivariate distribution $b_{12}(h_1, h_2)$ represents the probability of observing an effective rainfall depth $h_1$ in catchment 1 and an effective rainfall depth $h_2$ in catchment 2 during joint events. The choice of $b_{12}$ is a key point of the analysis because the synchronous increments of discharge (which result from synchronous rainfall events) determine an increase of correlation between the streamflow records at the two outlets (equation (5)). Here, $b_{12}$ is assumed to be a bivariate exponential distribution in the form proposed by *Srikanth Iyer et al. [2001]*. This expression has exponential marginals and its parameters are particularly easy to interpret. Moreover, this bivariate PDF allows the description of any positive linear dependence between the variates yet being parsimonious in terms of the number of parameters. In this case, the analytical expression for the Laplace transform of $b_{12}$, $\hat{\hat{B}}_{12}(s_1, s_2)$, reads [*Srikanth Iyer et al. 2001*]:

$$\hat{\hat{B}}_{12}(s_1, s_2) = \left[\frac{1}{1 + \alpha_1^{12}\,(s_1 + s_2 a\,)}\right]\left[\left(1 - a\frac{\alpha_1^{12}}{\alpha_2^{12}}\right)\frac{1}{\alpha_2^{12} s_2 + 1} + a\frac{\alpha_1^{12}}{\alpha_2^{12}}\right] \tag{6}$$

where $\alpha_1^{12}$ and $\alpha_2^{12}$ are the marginal means (the average effective rainfall depths in each catchment produced by joint events) and $a$ is a positive scaling factor that relates the variates by means of the relation: $h_2 = a h_1 + Z$ ($Z$ is an independent auxiliary random variable, whose distribution modulates the correlation between the two variates). According to equation (6), the correlation between the joint depths $h_1$ and $h_2$ ($r_\alpha$) can be expressed as [*Srikanth Iyer et al., 2001*]:

$$r_\alpha = a\frac{\alpha_1^{12}}{\alpha_2^{12}} \tag{7}$$

The shape of the bivariate PDF is thus controlled by the parameters $\alpha_1^{12}$, $\alpha_2^{12}$ and $a$ (or, alternatively, by $\alpha_1^{12}$, $\alpha_2^{12}$ and $r_\alpha$). Substituting equation (6) into equation (5), and taking advantage of equation (7), the following expression is obtained:

$$\int_0^\infty \frac{\partial^2 \hat{\hat{B}}_{12}(k_1 s_1 e^{-k_1 t}, k_2 s_2 e^{-k_2 t})}{\partial s_1 \partial s_2}\bigg|_{\substack{s_1=0\\ s_2=0}} dt = \frac{\alpha_1^{12}\alpha_2^{12} k_1 k_2\,(1 + r_\alpha)}{k_1 + k_2} \tag{8}$$

which will be used to specify the numerator of equation (5).

The solution of equation (5) further requires to specify the overall marginal PDFs of all the effective rainfall depths in the two contributing catchments (including joint and disjoint events) or, alternatively,

the PDF of the intensity of the disjoint effective rain depths. For this purpose, exponential univariate distributions have been chosen to preserve consistency with the joint PDF $b_{12}$ given via equation (6). The same assumptions has been done by *Rodriguez-Iturbe et al. [1999], Laio et al. [2001], Porporato et al. [2004], Müller et al. [2014]* and *Dralle et al. [2016]*. Three different cases, characterized by a different degree of complexity, are presented and discussed in the following. Each of these alternatives can be more or less suited to practical applications depending on data availability and the regional climatology. The first case (Section 2.3.1) represents the simplest model, which assumes that joint and disjoint events are drawn from the same population (i.e. they are characterized by the same depth distribution). The other cases, instead, identify two families of events featured by different statistical properties: joint and disjoint events (Case 2, discussed in section 2.3.2) or joint and overall events (Case 3, discussed in section 2.3.3). These models are more complicated and require a larger number of parameters. Case 3, in particular, is consistent with a class of models used for the characterization of flow duration curves *[Botter et al., 2007a, 2013]*. Since there are no theoretical reasons for which one of these models should be preferred a priori, we suggest the model selection should be case specific and performance driven.

### 2.3.1 *Case 1*

In this section, the PDF of the overall effective rainfall depths within each catchment is assumed to be equal to the marginal of the bivariate PDF of effective rainfall depths during joint events, i.e.: $b_{1t} = b_1^{12}$ and $b_{2t} = b_2^{12}$ (which implies $\hat{B}_{1t}(s_1) = \hat{\hat{B}}_{12}(s_1, 0)$ and $\hat{B}_{2t}(s_2) = \hat{\hat{B}}_{12}(0, s_2)$). Hence, the PDFs describing all the effective rainfall depths within each catchment are exponential distributions with mean $\alpha_i^{12}$. The assumption implies that the distribution of effective rainfall depths within each catchment is the same for joint and disjoint events. The Laplace transform $\hat{B}_{it}(s_i)$ associated to the PDF of the total effective rainfall depths $b_{it}$ for catchment i=1,2 reads:

$$\hat{B}_{it}(s_i) = \frac{1}{1 + \alpha_i^{12} s_i} \tag{9}$$

which leads to:

$$\int_0^\infty \frac{\partial^2 \hat{B}_{it}(k_i s_i e^{-k_i t})}{\partial {s_i}^2} dt \Big|_{s_i=0} = k_i \left(\alpha_i^{12}\right)^2 \tag{10}$$

Combining the result of equation (8) and (10) in equation (5), after some algebra the correlation between $q_1$ and $q_2$ can be written as:

$$\rho_{12} = \frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}} \frac{1}{2}(1 + r_\alpha) \frac{2\sqrt{k_1 k_2}}{k_1 + k_2} \tag{11}$$

The structure of equation (11) effectively highlights how different physical processes involved in the underlying streamflow dynamics affect the spatial correlation of streamflows. Three main drivers can be identified: the frequency of effective rainfall; the intensity of effective rainfall; the catchment transport properties. The physical drivers are represented by the three factors $F_\lambda$, $F_\alpha$ and $F_k$ constituting equation (11). Each factor is discussed below.

- $F_\lambda = \frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}}$ (relative frequency of joint streamflow producing rainfall events). This factor represents the relative frequency of synchronous streamflow-producing events (scaled to the geometric mean of the total frequency of events in each catchment). Because $\lambda_{it} = \lambda_{12} + \lambda_i$, the term $F_\lambda$ tends to one when only synchronous events take place. On the other hand, when joint events are not observed in the considered pair of catchments ($\lambda_{12} = 0$), the correlation drops to zero and the streamflow dynamics are uncorrelated, regardless of the other landscape and climate properties.

- $F_\alpha = \frac{1}{2}(1 + r_\alpha)$ (the arithmetic mean between 1 and the correlation of the intensities of the effective rainfall depths during joint events). This factor entails the effect of the correlation between the effective rainfall depths in the two catchments during joint events. Equation (11) shows that the correlation coefficient of the entire streamflow timeseries is linearly dependent on the correlation between the joint intensities of the effective rainfall in the contributing catchments. However, because of the synchronicity of the joint events and the temporal autocorrelation of the hydrographs, the spatial correlation of streamflow does not drop to zero even for uncorrelated joint depths ($F_\alpha \to 0.5$ when $r_\alpha$ tends to zero). Instead, when the intensity of the joint effective events in the two catchments are highly correlated, $F_\alpha \approx 1$.

- $F_k = \frac{2\sqrt{k_1 k_2}}{k_1 + k_2}$ (ratio between the geometric mean and the arithmetic mean of the recession rates). The last factor accounts for the heterogeneity in the geomorphological and hydrogeological features of the catchments, which result in different timescales of the hydrologic response in the contributing catchments. Note that $F_k \leqslant 1$, with $F_k = 1$ only when the recession rates in the two catchments are equal.

### 2.3.2 *Case 2*

In this case it is assumed that the PDF of the total intensity of the effective rainfall depths is a linear combination of two exponential distributions. The first describes the effective rainfall depths during

joint events, while the second refers to the depths of disjoint events. The former is an exponential distribution with mean $\alpha_i^{12}$, while the latter is an exponential distribution with mean $\alpha_i$. The PDF for the total effective rainfall depths within each catchment, $b_{it}$, is a mixed exponential distribution ($b_{it} = \frac{\lambda_i}{\lambda_{it}} b_i + \frac{\lambda_{12}}{\lambda_{it}} b_i^{12}$). Accordingly, the Laplace transform $\hat{B}_{it}(s_i)$ of $b_{it}$ can be written as:

$$\hat{B}_{it}(s_i) = \frac{\lambda_i}{\lambda_{it}} \left( \frac{1}{1+\alpha_i s_i} \right) + \frac{\lambda_{12}}{\lambda_{it}} \left( \frac{1}{1+\alpha_i^{12} s_i} \right) \tag{12}$$

which leads to:

$$\int_0^\infty \frac{\partial^2 \hat{B}_{it}(k_i s_i e^{-k_i t})}{\partial {s_i}^2} dt \Big|_{s_i=0} = k_i \frac{\lambda_i \left(\alpha_i\right)^2 + \lambda_{12} \left(\alpha_i^{12}\right)^2}{\lambda_{it}} \tag{13}$$

Inserting equations (8) and (13) into equation (5), the following expression for the spatial correlation of streamflow is obtained:

$$\rho_{12} = \frac{\lambda_{12} \alpha_1^{12} \alpha_2^{12}}{\sqrt{\left[\lambda_1 \left(\alpha_1\right)^2 + \lambda_{12} \left(\alpha_1^{12}\right)^2\right] \left[\lambda_2 \left(\alpha_2\right)^2 + \lambda_{12} \left(\alpha_2^{12}\right)^2\right]}} \frac{1}{2} \left(1 + r_\alpha\right) \frac{2\sqrt{k_1 k_2}}{k_1 + k_2} \tag{14}$$

The structure of the above solution is analogous to that of equation (11). However, in equation (14) the frequency and intensity of the effective rainfall are merged in the term

$$F_\lambda^{(2)} = \frac{\lambda_{12} \alpha_1^{12} \alpha_2^{12}}{\sqrt{\left[\lambda_1 \left(\alpha_1\right)^2 + \lambda_{12} \left(\alpha_1^{12}\right)^2\right] \left[\lambda_2 \left(\alpha_2\right)^2 + \lambda_{12} \left(\alpha_2^{12}\right)^2\right]}} \tag{15}$$

as a consequence of having calculated the overall distributions of the effective rainfall depth based on the relative frequency and intensity of joint and disjoint events (equation (12)). Therefore, in equation (14) the frequencies of the events are weighted based on their mean intensity. Note that, since $\lambda_{it} = \lambda_i + \lambda_{12}$, if $\alpha_i = \alpha_i^{12}$, equation (14) turns into equation (11). In fact, equation (11) can be interpreted as a special case of equation (14), corresponding to cases where the depth distributions of joint and disjoint events are equal.

### 2.3.3 *Case 3*

In this case it is assumed that, regardless of the shape of $b_{12}$, the overall distributions of the effective rainfall within each catchment ($b_{1t}$ and $b_{2t}$) are exponential distributions with means $\alpha_{1t}$ and $\alpha_{2t}$. In this case:

$$\hat{B}_{it}(s_i) = \frac{1}{1+\alpha_{it} s_i} \tag{16}$$

which leads to:

$$\int_0^{\infty} \frac{\partial^2 \hat{B}_{it}(s_i e^{-k_i t})}{\partial {s_i}^2} dt \Big|_{s_i=0} = k_i \alpha_{it}^2 \tag{17}$$

Combining equations (17), (8) and (5), the following expression for the streamflow correlation is finally obtained:

$$\rho_{12} = \frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}} \frac{\alpha_1^{12}\alpha_2^{12}}{\alpha_{1t}\alpha_{2t}} \frac{1}{2}(1+r_\alpha) \frac{2\sqrt{k_1 k_2}}{k_1+k_2} \tag{18}$$

Equation (18) is similar to equation (11) except for the presence of an additional term related to the effective rain depths ($\frac{\alpha_1^{12}\alpha_2^{12}}{\alpha_{1t}\alpha_{2t}}$). Such term quantifies the ratio between the mean intensities of the joint and the overall effective rainfall events. When $\alpha_i^{12} = \alpha_{it}$ (i.e. the mean depth of joint and total events is the same) equation (18) reduces to equation (11).

## 2.4 APPLICATION

As a proof of concept, we present here an application devoted to test the performances of the model in a real world setting. To this aim, equation (14) is selected because the statistical differences between the two classes of events (joint and disjoint) are accounted for more explicitly. In equation (14), in fact, the intensity and frequency of joint and disjoint events can be independently specified.

The performances of the analytical model are assessed by comparing the observed correlation of daily flows with the corresponding model estimate for a set of *16* catchments located in the Mid United States (Table 2). The study sites includes all the MOPEX catchments [*Schaake et al, 2006*] within a *120,000* km$^2$ (400 km x 300 km) region spanning across Arkansas, Missouri and Oklahoma (http://www.nws.noaa.gov/ohd/mopex/). All basins are weakly impacted by natural or artificial water storages (reservoirs, lakes) and are provided with daily streamflow records from *1948* to *2003*. The size of the catchments ranges from *400* to *7,500* km$^2$.

From the *16* case studies, *120* combinations of catchment pairs are obtained. The analysis is carried out at seasonal timescale, with seasons defined based on calendar dates (Spring: March, April, May; Summer: June, July, August; Autumn: September, October, November; Winter: December, January, February), leading to *480* couples of seasonal streamflow correlations. For each season, the measured streamflow (Pearson) correlation coefficient between the discharge timeseries observed at two arbitrary outlets ($\rho_{meas}$) is calculated as:

Table 2: Summary information about the study catchments

| Number | USGS Code | Name | Area [$km^2$] | Streamflow Station | State |
|---|---|---|---|---|---|
| 1 | 06928000 | Gasconade River | 3250 | Hazelgreen | MO |
| 2 | 06933500 | Gasconade River | 7384 | Jerome | MO |
| 3 | 07049000 | War Eagle Creek | 684 | Hindsville | AR |
| 4 | 07052500 | James River | 2566 | Galena | MO |
| 5 | 07056000 | Buffalo River | 2155 | St. Joe | AR |
| 6 | 07057500 | North Fork River | 1459 | Tecumseh | MO |
| 7 | 07058000 | Bryant Creek | 1482 | Tecumseh | MO |
| 8 | 07067000 | Current River | 4334 | Van Buren | MO |
| 9 | 07069500 | Spring River | 3068 | Imboden | AR |
| 10 | 07072000 | Eleven Point River | 2938 | Ravenden Springs | AR |
| 11 | 07074000 | Strawberry River | 1230 | Poughkeepsie | AR |
| 12 | 07186000 | Spring River | 3026 | Waco | MO |
| 13 | 07196500 | Illinois River | 2470 | Tahlequah | OK |
| 14 | 07197000 | Baron Fork | 811 | Eldon | OK |
| 15 | 07252000 | Mulberry River | 970 | Mulberry | AR |
| 16 | 07261000 | Cadron Creek | 434 | Guy | AR |

$$\rho_{meas} = \frac{\sum_{i=1}^{n} \left[ \left( q_1(i) - \langle q_1 \rangle \right) \left( q_2(i) - \langle q_2 \rangle \right) \right]}{\sqrt{\sum_{i=1}^{n} \left( q_1(i) - \langle q_1 \rangle \right)^2 \sum_{i=1}^{n} \left( q_2(i) - \langle q_2 \rangle \right)^2}} \tag{19}$$

where $q_1(i)$, and $q_2(i)$ are the streamflow at the outlet of the catchments 1 and 2 during the i-th day, $n$ is the number of recorded days, and $\langle q_1 \rangle$ and $\langle q_2 \rangle$ are the sample averages of $q_1$ and $q_2$.

The model parameters in equation (14) are estimated at seasonal timescale based on observed discharge timeseries, as discussed below. 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 $\lambda_i$ and $\lambda_{12}$ can therefore be inferred by counting the observed number of jumps in the daily streamflow records at the relevant outlets. These jumps are then classified as "disjoint" or "joint" according to their timing (joint events correspond to synchronous jumps in both catchments). The frequency of joint and disjoint events is then calculated by dividing the number of recorded events for the duration of the considered time series. Similarly, the average effective rainfall intensity, $\alpha$, can be evaluated from the magnitude of the daily streamflow jumps. 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 equation (1)). Consequently, $\alpha = \frac{<\Delta Q_j>}{k}$. The analysis is carried out for the different set of streamflow-producing events (joint and disjoint), thereby

allowing for the estimate of the corresponding mean depths ($\alpha_i$ and $\alpha_i^{12}$ respectively). In addition, $r_\alpha$ is estimated as the Pearson correlation coefficient between the joint streamflow increments in the two catchments (which are estimated as discussed above). Finally, the recession rate $k$ is evaluated from the observed hydrographs. Since we assume exponential recessions, the drainage rate $k$ is estimated by fitting a linear regression on different pairs $(\frac{\Delta Q}{\Delta t}, Q)$ selected from the descending limbs of observed hydrograph. [*Ceola et al., 2010; Basso et al., 2015; Dralle et al., 2015*].

The model succeeds in reproducing the observed variability of the seasonal streamflow correlation (Figure 3) among the study sites. The scatter plot of Figure 3 shows a good alignment along the 45° line, with a root mean square error *RMSE=0.086* and a mean absolute error *MAE=0.065*. A slight underestimation of the correlation is observed in Summer, when reduced discharges are likely to result in less robust estimates of the model parameters.

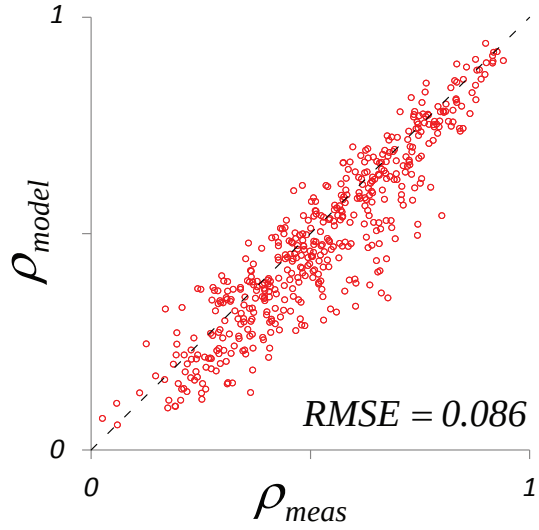


Figure 3: The scatterplot shows the performances of the model by comparing the observed and the modeled daily streamflow correlation between all the possible couples of catchment outlets within the study region. The application is performed at seasonal timescale: a single dot represents two catchments during a given season.

## 2.5 EFFECT OF SPATIALLY HETEROGENEOUS HYDROLOGICAL PROPERTIES ON THE STREAMFLOW CORRELATION

This section analyzes how the spatial correlation of streamflows is affected by the various parameters involved in the analytical formulation. For this purpose, the solution given via equation (14) — which also includes equation (11) as a special case — is considered. Nev-

ertheless, analogous results can be obtained using the other solution (equations (18)).

The results are discussed in terms of a synthetic dimensionless index, $V \in [0, 1]$, which expresses the inter-catchment variability of climatic and hydrologic properties. For each parameter involved in the analytical formulation we define the corresponding heterogeneity index as $V(*) = \frac{|*_1 - *_2|}{*_1 + *_2}$, where $*$ represents one of the parameters $k$, $\alpha^{12}$ or $\lambda_t$, and the subscripts 1 and 2 identify the relevant catchments. $V = 0$ implies spatial homogeneity of the considered attribute, while $V = 1$ implies enhanced heterogeneity of the underlying climate/-landscape properties.

The analysis is carried out with reference to the three factors of equation (14), namely $F_\lambda^{(2)}$, $F_\alpha$ and $F_k$. The impact of the heterogeneity of the frequency and intensity of effective rainfall events on the flow correlation is the most difficult to interpret because of the complex structure of $F_\lambda^{(2)}$. Hence, as a first approximation, we shall consider the special case where $\alpha_i = \alpha_i^{12}$, for which $F_\lambda^{(2)} = F_\lambda$ (equation (11)).

The upper panels of Figure 4 show how the heterogeneity in the frequency of runoff producing events affects the streamflow correlation through the factor $F_\lambda$. The plot shows that such an effect is strongly modulated by the frequency of the joint events $\lambda_{12}$. The decrease of correlation due to the heterogeneity in the frequencies of the overall effective rainfall events $V(\lambda_t)$ is more pronounced for higher frequencies of the joint events. The plot also shows that if $\lambda_{12} < \lambda_m$ ($\lambda_m = \min\{\lambda_{1t}, \lambda_{2t}\}$) there is a potentially significant loss of correlation regardless of the heterogeneity of the overall frequency of effective rainfall events (i.e. for $V(\lambda_t) = 0$). Moreover, an additional loss of correlation is observed when $\lambda_{1t} \neq \lambda_{2t}$, which is modulated by the magnitude of $V(\lambda_t)$. It is worth noting that the nested (or non-nested) nature of the catchments directly affects the spatial correlation of flows through the frequency factor $F_\lambda$. In particular, for nested catchments the frequency of joint events shall be equal to the total frequency of the effective rainfall events in the inner catchment (see Figure 5). In fact, the runoff produced from any nested sub-catchment of a river propagates along the network and affects the streamflow dynamics downstream, thereby implying that $\lambda_{12} = \lambda_m$ (left panel of Figure 5). Consequently, the case of nested catchments is described in Figures 4a and 4b by the upper blue lines, which imply larger values of $F_\lambda$ and $\rho_{12}$ for a given value of $V(\lambda_t)$, $V(k)$, $V(\alpha^{12})$ and $a$. Therefore, for a given degree of heterogeneity of climate and landscape attributes, the maximum streamflow correlation is achieved for nested catchments. Conversely, since non-nested catchments are flow-disconnected (right panel of Figure 5), the frequency of joint runoff events is lower than the minimum frequency of events in each basin. The only exception is the degenerate case where $\lambda_{12} = \lambda_{1t} = \lambda_{2t}$ and $F_\lambda = 1$. The same considerations drawn above concerning the effect

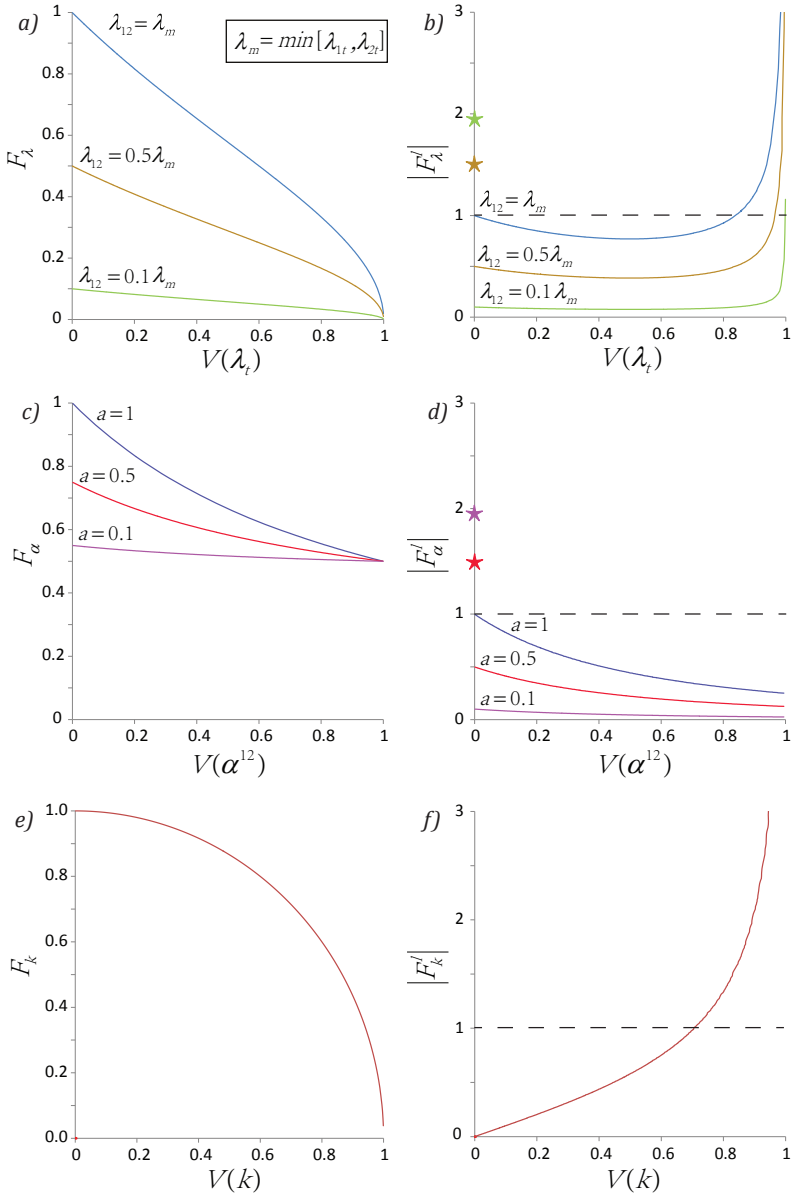


Figure 4: The plots show how the three factors in the analytical formulation change as a function of the heterogeneity of the physical parameters involved. The parameters include: i) the average effective rainfall frequencies ($\lambda_{it}$, $\lambda_{12}$); ii) the average joint effective rainfall depths ($\alpha_i^{12}$) and the correlation $r_\alpha$ between the joint effective rainfall depths (expressed by $a = r_\alpha \frac{\alpha_2^{12}}{\alpha_1^{12}}$); iii) the streamflow recession rates ($k_i$). The derivative of each factor (right panels) highlights the sensitivity of $\rho_{12}$ to the heterogeneity of these parameters. In panels 4c and 4d it is assumed $\alpha_2^{12} > \alpha_1^{12}$. The stars in Figure 4b and 4d indicate step-changes in $F_\lambda$ and $F_\alpha$ (and the y coordinate of each star indicates the relative extent of the corresponding step change).

of the variability in the frequency of effective rainfall events on $\rho_{12}$ hold in the general case $\alpha_i \neq \alpha_i^{12}$. However, in the latter case there is a significant impact of the ratios $\frac{\alpha_i}{\alpha_i^{12}}$ on the dependence between $\rho_{12}$ and $V(\lambda_t)$ (see Appendix 2.9).

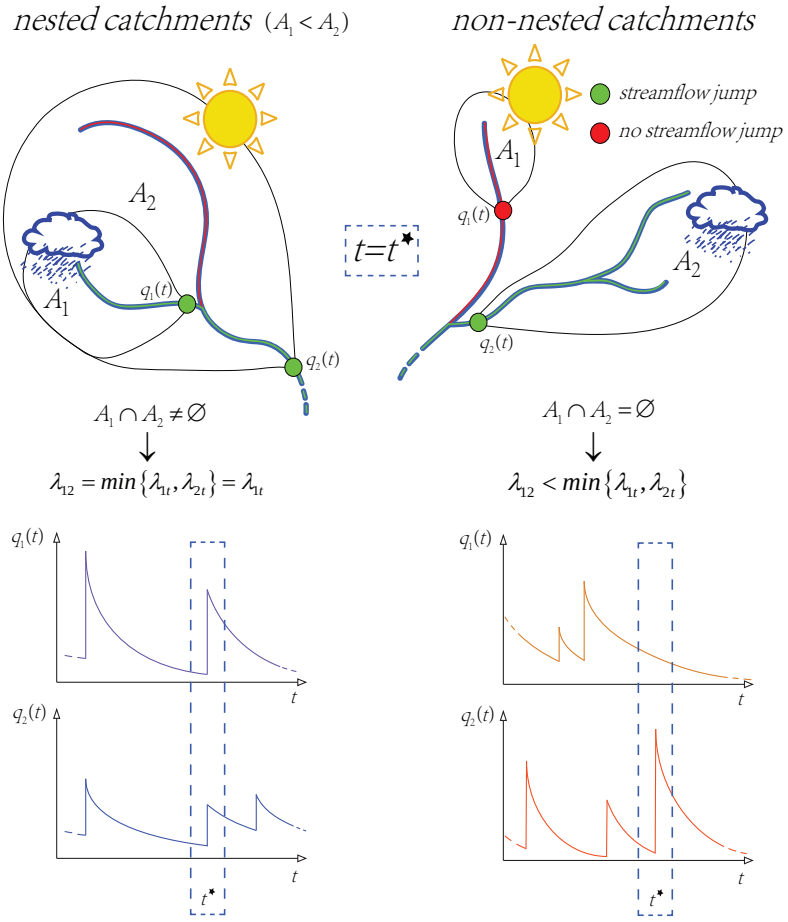


Figure 5: Implications of catchment arrangement within a river network for the flow correlation. Two configurations are possible: i) nested catchments: a smaller catchment ($A_1$) is nested within a larger one ($A_2$); ii) non nested catchments: the two catchment areas do not overlap. Due to downstream propagation of streamflows, in case of nested catchments, the frequency of joint runoff events equals the frequency of runoff events in the smaller catchment ($\lambda_{12} = \lambda_{1t}$). In case of non-nested catchments the frequency of joint events is usually smaller than the minimum runoff frequency in the two catchment ($\lambda_{12} < \min\{\lambda_{1t}, \lambda_{2t}\}$).

The second factor of equation (14), $F_\alpha = \frac{1}{2}(1 + r_\alpha)$, refers to the correlation between the depths of the joint streamflow producing events. In the plots of Figures 4c and 4d the correlation is expressed as a function of $a$ and $V(\alpha^{12})$, as $r_\alpha = a\left(\frac{2}{1-V(\alpha^{12})} - 1\right)^{-1}$. As expected, the correlation $\rho_{12}$ decreases with increasing $V(\alpha^{12})$. Though, Figure 4c and 4d show that the sensitivity to $V(\alpha^{12})$ decreases for low values

of $a$ and for large values of $V(\alpha_{12})$. Low values of $a$ imply a reduced proportionality between the magnitude of the effective rainfall depths in the two catchments. Hence, for low values of $a$, increasing $V(\alpha^{12})$ does not produce significant impacts on the streamflow correlation.

The factor $F_k = \frac{2\sqrt{k_1 k_2}}{k_1+k_2}$ in equation (14) describes how heterogeneity of the response times in the two contributing catchments influence the streamflow correlation at the corresponding outlets. Figure 4e and 4f highlight that the dependence of $F_k$ on $V(k)$ is described by the equation of a circle ($F_k$ is the ratio between the geometric and the arithmetic mean of the recession parameters). Hence, the correlation coefficient is almost insensitive to small and moderate heterogeneity in the transport properties of the catchments. The impact becomes considerable only for $V(k) \gtrsim 0.6$, which implies that $k_1$ and $k_2$ differ by almost an order of magnitude.

Overall, the analysis suggests that $F_\lambda$ and $F_k$ can independently drive the streamflow correlation to zero. However, the decrease of correlation due to heterogeneity in the frequency of effective rainfall is much larger than that due to differences in the hydrologic response, particularly when the frequency of joint events is relatively small and heterogeneity in recession rates is not exaggerated. Conversely, the heterogeneity in the joint effective rainfall depths described by $F_\alpha$ can decrease correlation by at most a factor of $\frac{1}{2}$. Nevertheless, the influence of the relative intensities of joint and disjoint events on the streamflow correlation should not be underestimated, as discussed in Appendix 2.9.

## 2.6 DISCUSSION

Despite the simplifications introduced to derive closed-form expressions of the seasonal streamflow correlation, the minimalist model proposed in this paper provides a formal linkage between the spatial correlation of daily flows and the underlying heterogeneity of climate and landscape features. Such a link helps to identify the hierarchy of physical controls acting on the spatial variability of flow regimes, which include the inter-catchment variability of precipitation regime, land cover and recession rates, as well as the topological arrangement of the contributing catchments.

In particular, the theoretical analysis points to the strong influence on the streamflow correlation played by the relative frequency ($\lambda_{12}$) and intensity ($\alpha_i^{12}$) of synchronous effective rainfall events. The analytical model suggests that the occurrence of intense joint flow-producing events in the selected catchments is a major driver of high correlation of streamflow timeseries. Vice versa: when the frequency of shared events between catchments is low (or the joint events are remarkably less intense than disjoint events), correlation between catchment outflows is also low. Note that the frequency of joint effective

rainfall $\lambda_{12}$ encapsulates different climate, soil and vegetation properties (and their inter-catchment variability) according to complex and strongly non-linear relationships [*Porporato et al., 2004; Isham et al., 2005; Botter et al., 2007a; Doulatyari et al., 2014*]. In fact, effective rainfall events represent precipitation events that fill the catchment-scale soil moisture deficit created by plant transpiration in the root zone [*Milly, 1994; Rodriguez-Iturbe et al., 1999; Laio et al., 2001*]. As such, the effective rainfall frequency is the byproduct of intertwined climate, soil and vegetation attributes (e.g. rainfall rates, soil storage capacity, dryness index). Therefore, the presence of synchronous events in the discharge time series at two selected outlets is influenced by two types of factors: i) the occurrence of joint rainfall events simultaneously feeding the relevant contributing catchments; ii) inter-catchment similarities of climate/landscape properties like the root zone depth and evapotranspiration rates. In fact, the presence of joint rainfall events is a necessary (though not sufficient) condition to observe high joint frequencies $\lambda_{12}$ (and thus high streamflow correlations). Similarities of landscape attributes also tends to increase the relative frequency of joint flow producing events in the two catchments because of the ensuing similarity in the soil moisture dynamics therein (which implies that the exceedance of the field capacity in the two catchments is more likely triggered by the same rainfall events).

Moreover, it is worth noting that spatial heterogeneity in evapotranspiration and soil properties bears a simultaneous impact both on the frequency of flow-producing events within each catchments ($\lambda_{it}$) and on the frequency of joint events ($\lambda_{12}$), with a potentially limited impact on the factor $F_\lambda = \frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}}$ (because of the simultaneous increase/decrease of the total and of the joint frequencies). Instead, since the frequency of joint rainfall events represents a physical upper bound for $\lambda_{12}$, heterogeneity in the rainfall forcing are more directly transmitted to the flow correlation. This instance seems to be an attractive feature of the proposed model, especially in view of possible application to ungauged sites, where $F_\lambda$ could be estimated from spatially interpolated daily precipitation records. As per the depths of effective rainfall, it has been shown elsewhere (e.g. *Rodriguez-Iturbe and Porporato [2005]*) that the main consequence of the interaction between rainfall and soil moisture dynamics is a decrease of the frequency of runoff events, with a more limited impact on the mean depth, at least for exponentially distributed rainfall depths [*Laio et al., 2001; Verma et al., 2011*].

Overall, the relative contribution of different climatic and landscape attributes to the streamflow correlation may be dependent on the spatial scales involved in the analysis (e.g. relative distance between catchments, size of the contributing areas) and on the specific climatic setting (e.g. correlation scale of rainfall properties). In small catchments, where climatic features can be expected to be relatively

constant over space, small-scale heterogeneity in geological properties and/or land cover (e.g. presence of karst areas or impervious regions) could significantly enhance the effect of spatially heterogeneous recession rates on the correlation of streamflows. Conversely, in larger catchments located in regions featured by strong climatic gradients (e.g. Alpine or pre-Alpine catchments where orographic effects may enhance the heterogeneity of rainfall) the streamflow correlation should be more strongly related to the spatial variability of rainfall and evapotranspiration. As expected, the topological arrangement of the two outlets also represents a major driver of the spatial correlation of daily flows. In case of nested catchments the correlation tends to increase as a byproduct of the following two combined agents: i) the frequency of joint events equals the minimum frequency of effective rainfall events in the two catchments (i.e. the inner catchment), which increases $\rho_{12}$ as discussed in Section 4; ii) the inter-catchment variability of climate, soil and vegetation properties is reduced (i.e. low values of $V$) because the two relevant contributing areas share a common region of the landscape. This latter effect should be particularly relevant when the two catchments have a similar size.

## 2.7 CONCLUSIONS

In this work we have derived a set of novel analytical expressions for the steady state linear correlation of daily discharges in two arbitrary locations of a river basin at seasonal time scale. The analytical development is based on the assumptions of Poisson effective rainfall and exponential recessions. The time lag between peak hydrographs due to flood wave propagation along the river network is also neglected, an instance which may prevent the application of the method to large ($A > 10^4$ km$^2$ ) basins. The resulting expressions for the streamflow correlation involve a limited number of hydrologic parameters that encapsulate soil/vegetation properties, precipitation regime and recession rates, and correspond to different assumptions on the distribution of effective rainfall depths.

The framework helps to identify the hierarchy of physical controls on the spatial variability of flow regimes. In particular, our theoretical analysis suggests that frequency and intensity of synchronous effective rainfall events in the relevant contributing catchments are the main drivers of the spatial correlation of daily flows, unless the heterogeneity of drainage rates is remarkable. As expected, topological arrangement of the considered outlets also influences the underlying correlation of daily flows. In fact, for nested catchments, the frequency of joint events is equal to the frequency of effective rainfall in the smaller catchment, which implies the maximization of the spatial correlation of discharge for a given degree of heterogeneity of climate and landscape properties in the two watersheds.

Model performances have been assessed by means of the application of the method to 16 catchments with a maximum size of 7,500 $km^2$ located in a 120,000 $km^2$ region in the United States. The application demonstrated the ability of the model to reproduce the observed streamflow correlations within the study region. Alternative procedures for the estimate of model parameters as well as the comparison among the different solutions in settings where different hydrological data are available are deferred to subsequent studies.

The proposed framework offers the opportunity to improve the characterization of the spatial and temporal variability of flow regimes within and across river basins and it may facilitate the prediction of flow regimes in poorly gauged areas or under changing climate conditions, with implications for water resources assessment and ecological studies.

## 2.8 APPENDIX

### 2.8.1 *Analytical derivation of the streamflow correlation*

The spatial correlation of the streamflows at the outlet of two catchments is defined as

$$\rho_{12} = \frac{cov(q_1, q_2)}{\sqrt{var(q_1)var(q_2)}} \tag{20}$$

where $cov(q_1, q_2) = \langle q_1 q_2 \rangle - \langle q_1 \rangle \langle q_2 \rangle$ is the streamflows covariance and $\langle q_i \rangle$ indicates the mean flow in catchment $i$. The expectation of the product of the streamfows $\langle q_1 q_2 \rangle$ can be obtained from the moment generating function (equation (4)), as:

$$\begin{aligned}\langle q_1 q_2 \rangle = \frac{d^2\hat{p}(s_1,s_2)}{ds_1 ds_2}\bigg|_{\substack{s_1=0\\s_2=0}} = \hat{p}(s_1,s_2)\bigg|_{\substack{s_1=0\\s_2=0}} \Bigg\{ \Bigg[\lambda_1 \int_0^\infty \frac{\partial \hat{B}_1(k_1 s_1 e^{-k_1 t})}{\partial s_1}\, dt + \\ + \lambda_{12} \int_0^\infty \frac{\partial \hat{\hat{B}}_{12}(k_1 s_1 e^{-k_1 t}, k_2 s_2 e^{-k_2 t})}{\partial s_1}\, dt\Bigg] \Bigg[\lambda_2 \int_0^\infty \frac{\partial \hat{B}_2(k_2 s_2 e^{-k_2 t})}{\partial s_2}\, dt + \\ + \lambda_{12} \int_0^\infty \frac{\partial \hat{\hat{B}}_{12}(k_1 s_1 e^{-k_1 t}, k_2 s_2 e^{-k_2 t})}{\partial s_2}\, dt\Bigg] + \lambda_{12} \int_0^\infty \frac{\partial^2 \hat{\hat{B}}_{12}(k_1 s_1 e^{-k_1 t}, k_2 s_2 e^{-k_2 t})}{\partial s_1 \partial s_2}\, dt \Bigg\}\Bigg|_{\substack{s_1=0\\s_2=0}}\end{aligned} \tag{21}$$

The first and the second factors between square brackets on the right hand side of equation (21) can be rearranged in terms of the Laplace transform of the 'total' probability density function of the jumps, $\hat{B}_{1t}$, which accounts for both joint and disjoint events. In fact, thanks to the independence of joint and disjoint events in equation

(2), the overall distribution of the effective rainfall depths, $b_{it}$, can be written as:

$$b_{it} = \frac{\lambda_i}{\lambda_{it}} b_i + \frac{\lambda_{12}}{\lambda_{it}} b_i^{12} \tag{22}$$

The Laplace transforms of equation (22) reads:

$$\hat{B}_{1t}(s_1) = \frac{\lambda_1}{\lambda_{1t}} \hat{B}_1(s_1) + \frac{\lambda_{12}}{\lambda_{1t}} \hat{\hat{B}}_{12}(s_1, 0) \;\; ; \;\; \hat{B}_{2t}(s_2) = \frac{\lambda_2}{\lambda_{2t}} \hat{B}_2(s_2) + \frac{\lambda_{12}}{\lambda_{2t}} \hat{\hat{B}}_{12}(0, s_2) \tag{23}$$

which allows equation (21) to be written in terms of $\hat{B}_{1t}$ as:

$$\begin{aligned} \langle q_1 q_2 \rangle \;\; = \;\; & \left[ \lambda_{1t} \int_0^\infty \frac{\partial \hat{B}_{1t}(k_1 s_1 e^{-k_1 t})}{\partial s_1} \, dt \right] \left[ \lambda_{2t} \int_0^\infty \frac{\partial \hat{B}_{2t}(k_2 s_2 e^{-k_2 t})}{\partial s_2} \, dt \right] + \\ + \;\; & \lambda_{12} \int_0^\infty \frac{\partial^2 \hat{\hat{B}}_{12}(k_1 s_1 e^{-k_1 t}, k_2 s_2 e^{-k_2 t})}{\partial s_1 \partial s_2} \, dt \Bigg|_{s_1=0, s_2=0} \end{aligned} \tag{24}$$

Similarly, the average streamflow at the outlet of the i-th catchment, $\langle q_i \rangle$, can be obtained from the moment generating function of the corresponding marginal streamflow distributions, which can be expressed as [*Kingman, 1992*]:

$$\hat{p}(s_i) = \exp\left\{ -\lambda_{it} \int_0^\infty \left[ 1 - \hat{B}_{it}(k_i s_i e^{-k_i t}) \right] dt \right\} \qquad (i = 1, 2) \tag{25}$$

Equation (25) leads to the following expression of the expected discharge at the outlet of catchment $i$:

$$\langle q_i \rangle = -\frac{\partial \hat{p}(s_i)}{\partial s_i} \Bigg|_{s_i=0} = -\lambda_{it} \int_0^\infty \frac{\partial \hat{B}_{it}(k_i s_i e^{-k_i t})}{\partial s_i} dt \Bigg|_{s_i=0} \qquad (i = 1, 2) \tag{26}$$

Combining equation (24) and (26), the covariance between the streamflow at the two catchments can then be expressed as:

$$cov(q_1, q_2) = \lambda_{12} \int_0^\infty \frac{\partial^2 \hat{\hat{B}}_{12}(k_1 s_1 e^{-k_1 t}, k_2 s_2 e^{-k_2 t})}{\partial s_1 \partial s_2} dt \Bigg|_{s_1=s_2=0} \tag{27}$$

Equation (27) states that only joint effective rainfall events generate positive correlations between $q_1$ and $q_2$.

To obtain the final expression of the correlation, the covariance needs to be normalized by means of the product of the standard deviations. To this aim, the streamflow variance at each outlet, $var(q_i)$, is expressed using the moment generating function (equation (25)) as:

$$\begin{aligned} var(q_i) &= \langle q_i^2 \rangle - \langle q_i \rangle^2 = \frac{\partial^2 \hat{p}(s_i)}{\partial {s_i}^2}\bigg|_{s_i=0} - \left[\lambda_{it} \int_0^\infty \frac{\partial \hat{B}_{it}(k_i s_i e^{-k_i t})}{\partial s_i} dt\right]^2\bigg|_{s_i=0} = \\ &= \lambda_{it} \int_0^\infty \frac{\partial^2 \hat{B}_{it}(k_i s_i e^{-k_i t})}{\partial {s_i}^2} dt\bigg|_{s_i=0} \end{aligned} \tag{28}$$

Finally, combining equation (20), (27) and (28), the spatial correlation between the streamflows $q_1$ and $q_2$ can be written as in equation (5) of the main text.

## 2.9 EFFECT OF HETEROGENEOUS MEAN DEPTHS OF JOINT-DISJOINT EFFECTIVE RAINFALL EVENTS ON $\rho_{12}$

The effect on $\rho_{12}$ due to the frequencies of joint and disjoint events is here assessed in the general case of $\alpha_i \neq \alpha_i^{12}$. In this case, in the factor $F_\lambda^{(2)}$ the influence of the frequency of joint and disjoint effective rainfall events is weighted based on their correspondent mean depths.

Figure 6 shows how $F_\lambda^{(2)}$ (and hence the correlation $\rho_{12}$), decreases as a function of $V(\lambda_t)$ for different frequencies $\lambda_{12}$ and different combinations of the spatial heterogeneity between the mean intensity of joint and disjoint events. The latter is quantified by means of ratios $\delta_i = \frac{\alpha_i}{\alpha_i^{12}}$. Figure 6 shows how the ratio $\delta_1$ modulates the dependence of $F_\lambda^{(2)}$ on $V(\lambda_t)$ when $\lambda_{12} = \lambda_m$ (which implies $\lambda_2 = 0$). When $\delta_1 < 1$ (high relative intensity of joint events), the effect of joint effective rainfalls is enhanced and high values of $F_\lambda^{(2)}$ are maintained for a wide range of $V(\lambda_t)$. Conversely, high values of $\delta_1$ ensue a faster loss of correlation, despite relatively high frequency of joint events (low values of $V(\lambda_t)$).

In Figure 6 the effect of $\delta_i$ is assessed in the case $\lambda_{12} = 0.5\,\lambda_m$. In this case the heterogeneity of the intensities affects $F_\lambda^{(2)}$ also when $V(\lambda_t) = 0$. Higher values of correlation are ensured by low values of $\delta_i$. On the contrary, a significant drop of correlation is observed when the intensities of disjoint events are high compared to the intensities of joint events (i.e. for higher values of $\delta_i$).

This analysis pinpoints the intertwined role of the frequency and intensity of effective rainfall events on the streamflow correlation. In particular, heterogeneity in the relative mean depths between joint and disjoint effective rainfall can strongly impact the dependence of $F_\lambda^{(2)}$ on $V(\lambda_t)$.

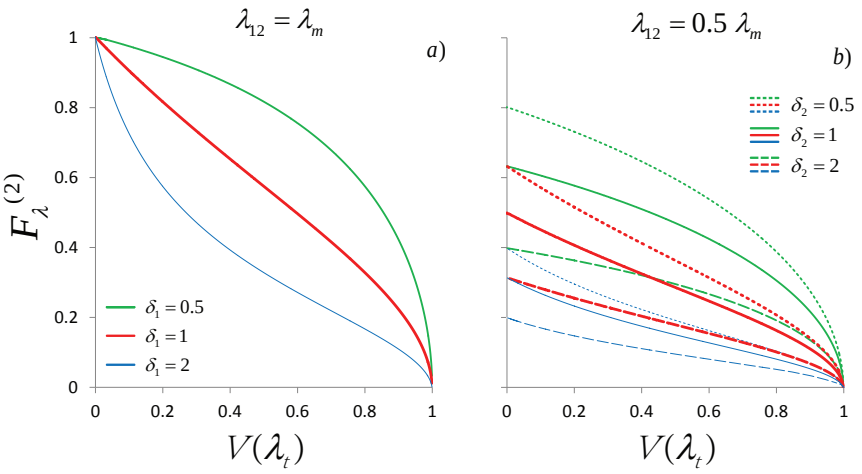


Figure 6: Effect of heterogeneous joint and disjoint effective rainfall intensities on streamflow correlation for different frequencies of effective rainfall. Differences between joint and disjoint effective rainfall intensities within the same catchment $i$ are quantified by the ratio $\delta_i = \frac{\alpha_i}{\alpha_i^{12}}$. The couple of ratios $(\delta_1, \delta_2)$ identify each curve. Without loss of generality it is assumed $\lambda_{1t} > \lambda_{2t}$. In panel $a)$ the curves are independent on the parameter $\delta_2$ (because $\lambda_2 = 0$). The continuous red curve corresponds to equation (11) and Figure 4. As defined in the text, $\lambda_m = \min\{\lambda_{1t}, \lambda_{2t}\}$.

## ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 641939. This study was also funded by the Swiss National Science Foundation (SNF, Project No.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 dataset and are freely available online at: http://www.nws.noaa.gov/ohd/mopex/.

The authors would like to thank the three anonymous reviewers and the associate editor for having contributed to improve the quality of this work with their comments.

# 3

# WHAT DO THEY HAVE IN COMMON? DRIVERS OF STREAMFLOW SPATIAL CORRELATION AND PREDICTION OF FLOW REGIMES IN UNGAUGED LOCATIONS

Andrea Betterle[1,2,3], Dirk Radny[1], Mario Schirmer[1,2], Gianluca Botter[3]
Water Resources Research, 2017

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

## 3.1 INTRODUCTION

The spatial and temporal variability of streamflows critically impact the life of riverine biomes and the security of anthropogenic water uses [*Postel and 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 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 and 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. *Schaefli et al., 2014, Doulatyari et al., 2017*]. 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 [*Messinger and Paybins, 2014, Chacon-Hurtado et al., 2017*] 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 flow correlation represents a better surrogate than spatial proximity to individuate river sections characterized by analogous flow dynamics [*Archfield and 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*]. 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 non-nested) nature of catchments is considered in the definition of model parameters [*Müller and*

*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 [*Skøjen and Blöschl, 2007; Blöschl et al., 2013; Müller and Thompson, 2015*]. The analytical structure of the model proposed by *Bettrele 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 inter-catchment 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 $75,000\,\mathrm{km}^2$ 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 3.2 summarizes the analytical framework used to perform the analysis; section 3.3 presents the case studies and the hydrologic data. Section 3.4 discusses two alternative approaches to estimate the parameters of the model. Sec-

tion 3.5 discusses the performances of the models and relevant results. Section 3.6 concludes the paper.

## 3.2 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 smaller than one 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, 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 [*Claps et al., 2005; Botter et al., 2007a; 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; Pumo et al., 2013; Müller and Thompson, 2015*]. 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 [*Milly, 1994; Porporato et al., 2004; Thompson et al., 2011; Botter et al., 2013*]. The effective rainfall within each contributing area is modeled according to a Poisson process of frequency $\lambda_{it}$, 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 $\alpha_{it}$.

The streamflow correlation between two selected outlets is then studied by decoupling the "total" sequence of effective rainfall events

Table 3: Summary of the parameters

| Parameter | Description |
|---|---|
| $\lambda_{it}$ | Average frequency of all effective rainfall events in catchment $i$ |
| $\lambda_i$ | Average frequency of disjoint effective rainfall events in catchment $i$ |
| $\lambda_{12}$ | Average frequency of joint effective rainfall events in the two catchments |
| $\lambda_m$ | Minimum between $\lambda_{1t}$ and $\lambda_{2t}$ |
| $\alpha_{it}$ | Average depth of all effective rainfall in catchment $i$ |
| $\alpha_i$ | Average depth of disjoint effective rainfall in catchment $i$ |
| $\alpha_i^{12}$ | Average depth of joint effective rainfalls 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 in catchment $i$ |
| $\tilde{\alpha}_i$ | Average depth of disjoint rainfall in catchment $i$ |
| $\tilde{\alpha}_i^{12}$ | Average depth of joint rainfalls in catchment $i$ |
| $r_\alpha$ | 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 $i$ |

i=1,2 identifies the two catchments;
joint: affects at the same time both catchments;
disjoint: affects only one catchment at a time.

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, $\lambda_{12}$ and $\lambda_i$ 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 $\alpha_i^{12}$ and $\alpha_i$, respectively. For a detailed summary of all model parameters the reader is directed to table 3.

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

$$\rho_{model}^{(1)} = \overbrace{\underbrace{\frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}}}_{F_\lambda^{(1)}} \underbrace{\frac{1}{2}(1+r_\alpha)}_{F_\alpha^{(1)}}}^{F_{\lambda\alpha}^{(1)}} \underbrace{\frac{2\sqrt{k_1 k_2}}{k_1+k_2}}_{F_k^{(1)}} \tag{29}$$

$$\rho^{(2)}_{model} = \overbrace{\frac{\lambda_{12}\alpha_1^{12}\alpha_2^{12}}{\sqrt{\left[\lambda_1(\alpha_1)^2+\lambda_{12}\left(\alpha_1^{12}\right)^2\right]\left[\lambda_2(\alpha_2)^2+\lambda_{12}\left(\alpha_2^{12}\right)^2\right]}}\frac{1}{2}(1+r_\alpha)}^{F^{(2)}_{\lambda\alpha}} \overbrace{\frac{2\sqrt{k_1k_2}}{k_1+k_2}}^{F^{(2)}_{k}} \tag{30}$$

$$\rho^{(3)}_{model} = \overbrace{\overbrace{\frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}}}^{F^{(3)}_{\lambda}} \overbrace{\frac{\alpha_1^{12}\alpha_2^{12}}{\alpha_{1t}\alpha_{2t}}\frac{1}{2}(1+r_\alpha)}^{F^{(3)}_{\alpha}}}^{F^{(3)}_{\lambda\alpha}} \overbrace{\frac{2\sqrt{k_1k_2}}{k_1+k_2}}^{F^{(3)}_{k}} \tag{31}$$

where '$r_\alpha$' 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 (29) 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 (30) assumes that joint and disjoint effective rainfall depths are characterized by two different exponential probability density functions (PDF) with means $\alpha_i^{12}$ and $\alpha_i$ respectively. Equation (31) assumes two independent exponential PDFs for joint and total effective rainfall depths, with means $\alpha_i^{12}$ and $\alpha_{it}$, respectively. Therefore, equations (30) and (31) 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 vs. local convective or orographic events). Equation (29) represents the simplest model, and it can be interpreted as a special case of equation (30) and (31) for $\alpha_i = \alpha_i^{12}$ and $\alpha_{it} = \alpha_i^{12}$ respectively.

Equations (29) to (31) can ideally be decomposed into a number of factors that identify the contribution to the streamflow correlation associated to different hydrological processes, namely: i) the relative frequency of joint effective rainfall events ($F_\lambda$); ii) the intensity of effective rainfall events ($F_\alpha$), and iii) the features of recession rates ($F_k$). Note that in equation (30) the effect of the frequency and intensity of events is blended into a single term, $F^{(2)}_{\lambda\alpha}$. The superscripts of the terms $F_\lambda$, $F_\alpha$, $F_k$ and $F_{\lambda\alpha}$ refer to the specific solution considered (equations (29) to (31)) .

Note that the model is able to account for the topological arrangement of the considered catchments, as the nested or non-nested nature of the two catchments is reflected in the frequency of joint effective rainfall events ($\lambda_{12}$). 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 sub-catchment of

a river indeed propagates along the network and affects the streamow dynamics in downstream sites. In case of non-nested catchments, on the other hand, the frequency of joint events is typically smaller than the minimum runoff frequency in the two catchments: $\lambda_{12} \leqslant \min\{\lambda_{1t}, \lambda_{2t}\}$. More details on the derivation of the model equations and hypothesis are given in *Betterle et al. [2017]*.

## 3.3 CASE STUDIES AND HYDROLOGIC DATA

The analytical model (equations (29) to (31)) is applied to *13* nested and non-nested catchments in a $75,000\ km^2$ region of the eastern US (Table 4). 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 7). 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 dataset (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 dataset are provided by USGS (https://waterdata.usgs.gov/nwis/rt), whereas spatially averaged daily rainfall rates are based on the PRISM model (http://www.prism.oregonstate.edu) [*Daly, 2008*]. Minimum and maximum average values of mean annual precipitation, mean annual streamflow, altitude and slope across the study catchments are respectively: $p_{min-max} = 3.0 - 5.4\ mm/day$, $q_{min-max} = 1.3 - 3.8\ mm/day$, $h_{min-max} = 480 - 1220\ m$, $s_{min-max} = 1.8\% - 6.4\%$. Land cover in the study area is rather heterogeneous and varies from evergreen needle leaf forest (0-50%) to deciduous broadleaf forests (20-90%), grassland and cropland (0-15%) (University of Maryland land use classification [*Hansen et al., 2000*]). Despite the pronounced inter-seasonal 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 was 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, Novem-

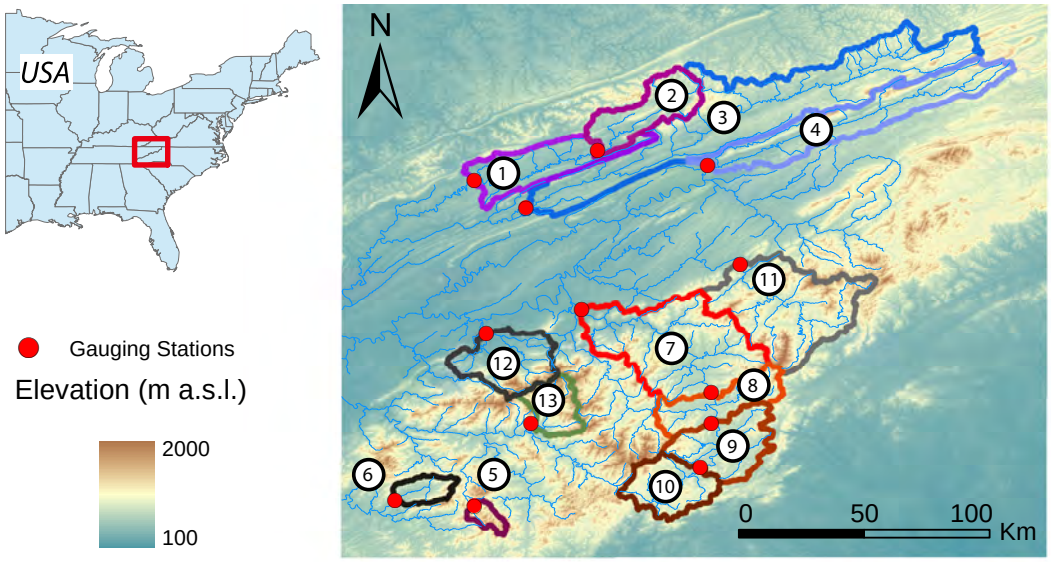


Figure 7: The study area includes 13 nested and non-nested catchments located in the eastern United States.

Table 4: Summary information about the study catchments

| Number | USGS Code | Name | Area [$km^2$] | Min - Max Altitude [m a.s.l.] | Streamflow Station | State |
|---|---|---|---|---|---|---|
| 1 | 03532000 | Powell River | 1774 | 328 - 1263 | Arthur | TN |
| 2 | 03531500 | Powell River | 826 | 394 - 1263 | Jonesville | VA |
| 3 | 03528000 | Clinch River | 3818 | 335 - 1424 | Tazewell | TN |
| 4 | 03490000 | Holston River | 1738 | 370 - 1442 | Gate City | VA |
| 5 | 03504000 | Nantahala River | 134 | 958 - 1643 | Rainbow Springs | NC |
| 6 | 03550000 | Valley River | 269 | 482 - 1532 | Tomotla | NC |
| 7 | 03455000 | French Broad River | 4812 | 317 - 1925 | Newport | TN |
| 8 | 03451500 | French Broad River | 2448 | 570 - 1925 | Asheville | NC |
| 9 | 03448000 | French Broad River | 1751 | 608 - 1808 | Bent Creek | NC |
| 10 | 03443000 | French Broad River | 767 | 634 - 1808 | Blantyre | NC |
| 11 | 03465500 | Nolichucky River | 2085 | 469 - 2002 | Embreeville | TN |
| 12 | 03470000 | Little Pigeon River | 914 | 273 - 1998 | Sevierville | TN |
| 13 | 03512000 | Oconaluftee River | 477 | 573 - 1877 | Birdtown | NC |

ber; 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 ($\rho_{meas}$) can be calculated as:

$$\rho_{meas} = \frac{\sum_{j=1}^{n} \left[ (q_1(j) - \langle q_1 \rangle)(q_2(j) - \langle q_2 \rangle) \right]}{\sqrt{\sum_{j=1}^{n} (q_1(j) - \langle q_1 \rangle)^2 \sum_{j=1}^{n} (q_2(j) - \langle q_2 \rangle)^2}} \tag{32}$$

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 (32) is then used to evaluate model performances as discussed in what follows.

## 3.4 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.

### 3.4.1 *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 ($\lambda_{it}$, $\lambda_i$ and $\lambda_{12}$) 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, $\alpha$, 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 ($\alpha_{it}$, $\alpha_i^{12}$ and $\alpha_i$, respectively). In addition, $r_\alpha$ 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 [*Ceola et al., 2010; Basso et al., 2015; 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 por-

tray the signature of catchment-scale soil moisture variability, which in turn encapsulate the effect of climatic and landscape attributes [e.g. *Milly, 1994; Rodriguez-Iturbe et al., 1999; Laio et al., 2001; 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 [*Chen and Krajewski, 2016*].

### 3.4.2 *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 inter-catchment variability of rainfall properties is likely to be a primary control on the spatial variability of the frequency and intensity of effective rainfall events. 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 and 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 $F_\lambda$, $F_\alpha$ and $F_{\lambda\alpha}$ in equations (29), (30) and (31), can be calculated based on rainfall data, instead of effective rainfall time series. More specifically, in equation (29) we assume that:

$$\frac{\lambda_{12}}{\sqrt{\lambda_{1t}\lambda_{2t}}} \approx \frac{\tilde{\lambda}_{12}}{\sqrt{\tilde{\lambda}_{1t}\tilde{\lambda}_{2t}}} \tag{33}$$

and in equation (30)

$$\frac{\lambda_{12}\alpha_1^{12}\alpha_2^{12}}{\sqrt{\left[\lambda_1(\alpha_1)^2+\lambda_{12}\left(\alpha_1^{12}\right)^2\right]\left[\lambda_2(\alpha_2)^2+\lambda_{12}\left(\alpha_2^{12}\right)^2\right]}} \approx$$
$$\approx \frac{\tilde{\lambda}_{12}\tilde{\alpha}_1^{12}\tilde{\alpha}_2^{12}}{\sqrt{\left[\tilde{\lambda}_1(\tilde{\alpha}_1)^2+\tilde{\lambda}_{12}\left(\tilde{\alpha}_1^{12}\right)^2\right]\left[\tilde{\lambda}_2(\tilde{\alpha}_2)^2+\tilde{\lambda}_{12}\left(\tilde{\alpha}_2^{12}\right)^2\right]}} \tag{34}$$

where the tilde (~) 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 i-th 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 non-zero 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 $r_\alpha$ between the intensity of joint effective rainfall time series in equations (29) to (31) is estimated as the correlation between the corresponding joint rainfall events $r_{\tilde{\alpha}}$. In the same vein, in equation (31), 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 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 [*Botter et al., 2007c; Pumo et al., 2013; Botter et al., 2013; Mejia et al., 2014; Müller et al., 2014; Doulatyari et al., 2015; Basso et al., 2015; Müller and Thompson, 2015*].

Finally, since the estimate of recession properties is challenging in the absence of discharge data [*Biswal and 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 $F_\lambda$, $F_\alpha$ and $F_{\lambda\alpha}$ (equations (29) to (31)), 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. $\lambda_{12}$) is facilitated by the use of synchronous rainfall records, which allows an easier estimate of the timing of rainy days and the corresponding rainfall depths.

Given the importance of an accurate estimate of the spatio-temporal gradients of rainfall, reliable information of 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 geo-statistical 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 [*Lai and Katul, 2000; Laio et al., 2001; Doulatyari et al., 2017*].

## 3.5 RESULTS AND DISCUSSION

### 3.5.1 *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 is compared to the estimates provided by the analytical model — equations (29) to (31) — using different parameter estimation procedures (sections 3.4.1 and 3.4.2). The distribution of the model parameters across the case studies is summarized in the boxplots in Figure 8. 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 8. As expected, the overall frequency of events ($\lambda_{it}$, $\tilde{\lambda}_{it}$) 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 8 also displays the distribution of the correlation coefficient between joint effective rainfall depths ($r_\alpha$) 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 inter-catchment 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 9 represents the model performance when parameters are estimated from discharge time series as discussed in section 3.4.1. Equations (30) (Fig. 9b) and equation (31) (Fig. 9c) capture reasonably well the observed variability of streamflow correlation across the study catchments. Conversely, equation (29) (Fig. 9a) systematically underestimates the observed correlations because it disregards

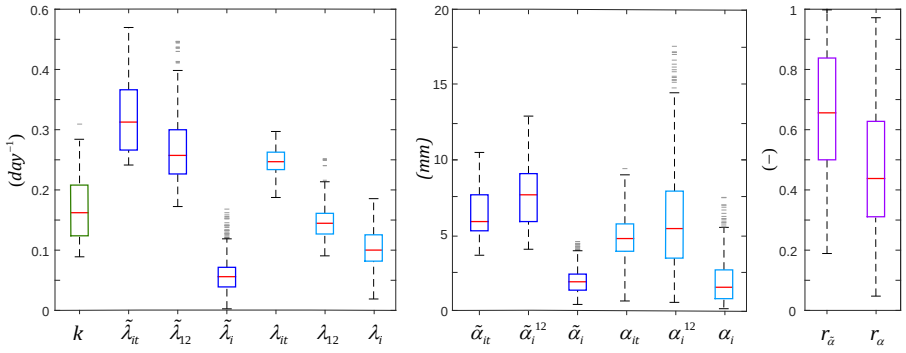


Figure 8: The boxplots on the left show the distribution of the streamflow recession rates (k), the frequency of total, joint and disjoint rainfall events ($\tilde{\lambda}_t$, $\tilde{\lambda}_{12}$, $\tilde{\lambda}_i$) and the frequency of total, joint and disjoint effective rainfall events ($\lambda_t$, $\lambda_{12}$, $\lambda_i$). 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 ($r_\alpha$) and the correlation of the joint rainfall depths ($r_{\tilde{\alpha}}$)

the difference between the intensity of joint and disjoint effective rainfall events. The underestimation is enhanced for intermediate values of $\rho_{meas}$, 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 $\rho^{(1)}_{model}$. On the contrary, $\rho^{(1)}_{model}$ 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 (30) is the best performer (RMSE $= 0.094$), followed by equation (31). Note that joint events can be difficult to individuate and quantify from streamflow time series, leading to underestimated values $\lambda_{12}$ or biased estimates of $\alpha_i^{12}$. This could explain the slight underestimation of $\rho^{(2)}_{model}$, 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 (30) further increases when the intensity of effective events is directly estimated from rainfall time series (Fig. 9d), with a RMSE equal to 0.056. Overall, the superior performances of equation (30) 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 10 shows the results obtained by means of equations (29), (30) and (31) when the parameters are estimated without exploiting discharge data, as discussed in section 3.4.2. As expected, a general decrease of performance is observed. In particular, the higher corre-

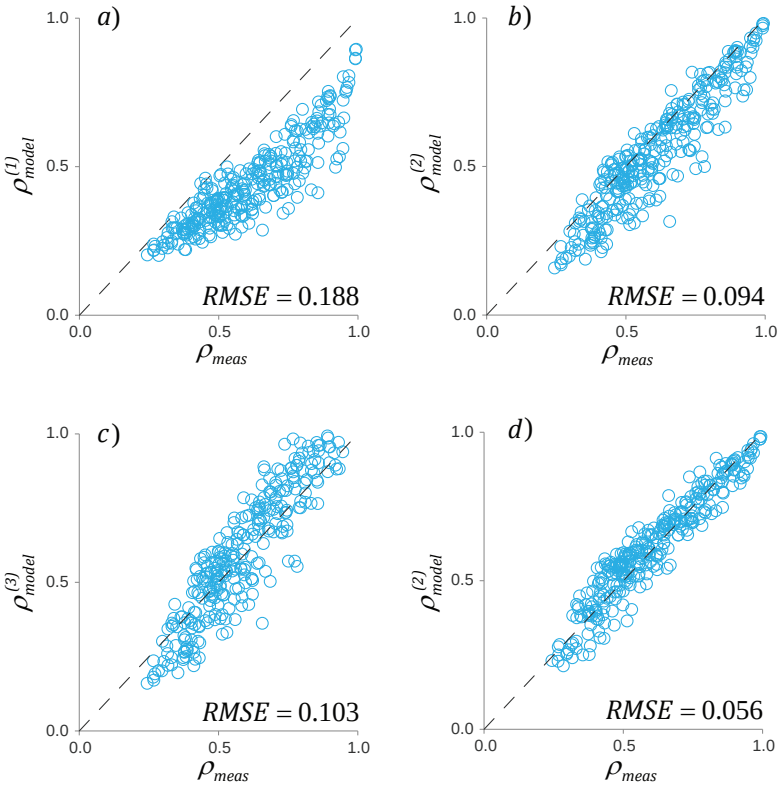


Figure 9: The scatterplots compare the seasonal streamflow correlation calculated from daily discharge records (x-axis) and the streamflow correlation estimated by the analytical model (equations (29) to (31)) when the parameters are obtained from streamflow records (y-axis) for the 312 pairs of seasonal streamflow time series between the 13 study catchments. Scatterlplots a), b) and c) refer to equations (29), (30) and (31) respectively. Scatterplot d) shows the performances of equation (30) when the intensities of joint effective rainfall events are estimated directly from rainfall records.

lation of joint rainfall events — if compared to the corresponding correlation of effective rainfall events, see Figure 8 — leads to overestimated streamflow correlations in all versions of the model. In this case, the simplest version of the model, represented by equation (29) (Figure 10a), benefits more than the others from the use of rainfall-estimated model parameters and outperforms equation (30) and (31). 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 10d shows the performances of equation (29) when the assumption $F_k = 1$ is relaxed, and $F_k$ is calculated from observed streamflow records as in Figure 9. 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.

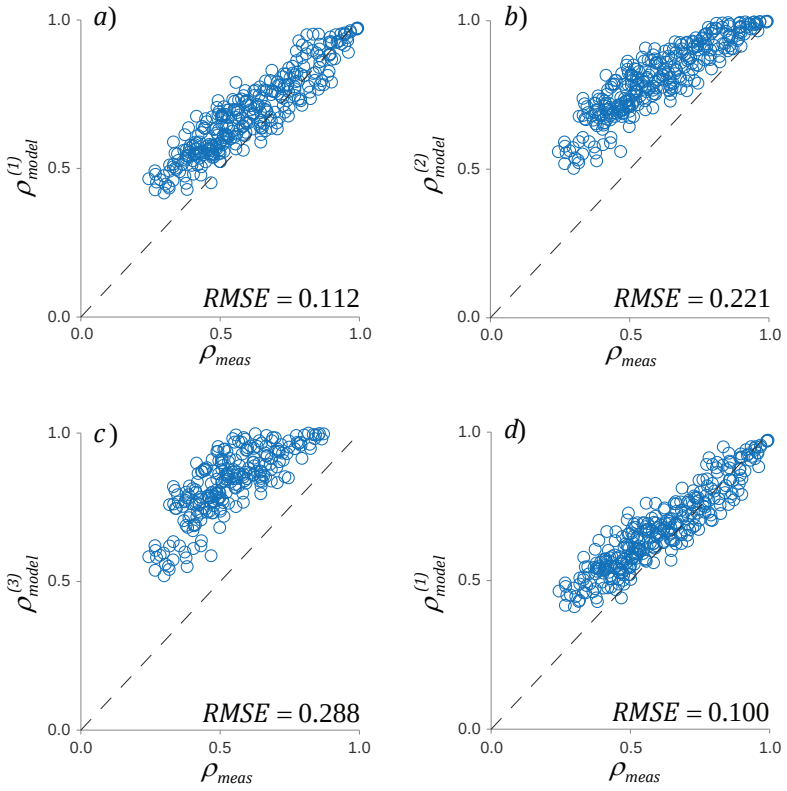


Figure 10: The scatterplots compare the streamflow correlation calculated from daily discharge records (x-axis) and the streamflow correlation estimated by the analytical model (equations (29) to (31)) when the parameters are obtained directly from rainfall (y-axis) for the 312 pairs of seasonal streamflow time series between the 13 study catchments. Scatterplots a), b) and c) refer to equation 29, 30 and 31 respectively assuming $F_k = 1$. Scatterplot d) shows the performances of equation (29) when the streamflow recession parameters are estimated based on discharge data.

The satisfactory performance provided by equation (29) 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 3.5.3). In particular, the successful application of equation (29) 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 streamflow characteristics at ungauged sites. The improved model performances when soil moisture dynamics are indirectly accounted for (parameters estimated based on streamflow records as in section 3.4.1 and

Figure 9) or when the assumption of homogeneous recession rates is relaxed (Figure 10d), 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 co-evolution of landscapes experiencing analogous climate and rainfall regimes [*Sivapalan, 2011; Jefferson, 2010; Huang 2006*], can justify the application of the framework in absence of discharge data. Especially when the identification of highly correlated sites is concerned.

Model performances are further analyzed in Figure 11, which displays the results obtained at annual timescale using equations (29), (30) and (31), when model parameters are estimated using different methods (see caption of Figure 11 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.

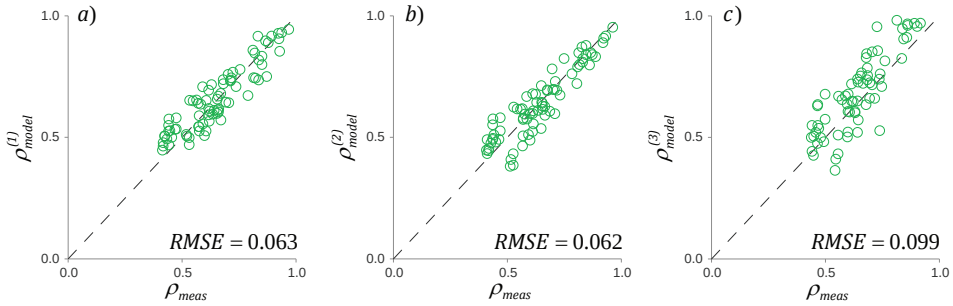


Figure 11: The scatterplots compare the streamflow correlation calculated from daily discharge records (x-axis) and the streamflow correlation estimated by the analytical model (y-axis) for the 78 pairs of annual streamflow time series selected from the 13 study catchments. In panel a) streamflow correlation is estimated by means of equation (29) when $F_\lambda$ and $F_\alpha$ are obtained from rainfall data and $F_k$ is estimated from discharge time series. In panel b) streamflow correlation is estimated by means of equation (30) when $F_\lambda$ and $F_k$ are obtained from streamflow records and $F_\alpha$ is obtained from rainfall records. In panel c) streamflow correlation is estimated by means of equation (31) when model parameters are obtained from streamflow records.

The plots in Figure 12 show the patterns of streamflow correlation across the study area as a function of the distance between the centre of mass of the contributing catchments. The observed trends are properly portrayed by the analytical formulation (in this case equation (30) is used with parameters estimated from streamflow data). Figure 12 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 heterogene-

ity of climatic and landscape properties in the study area. Scattering and seasonal variability of the relationship between the spatial correlation of discharges and distance seem to be adequately captured by the analytical model.

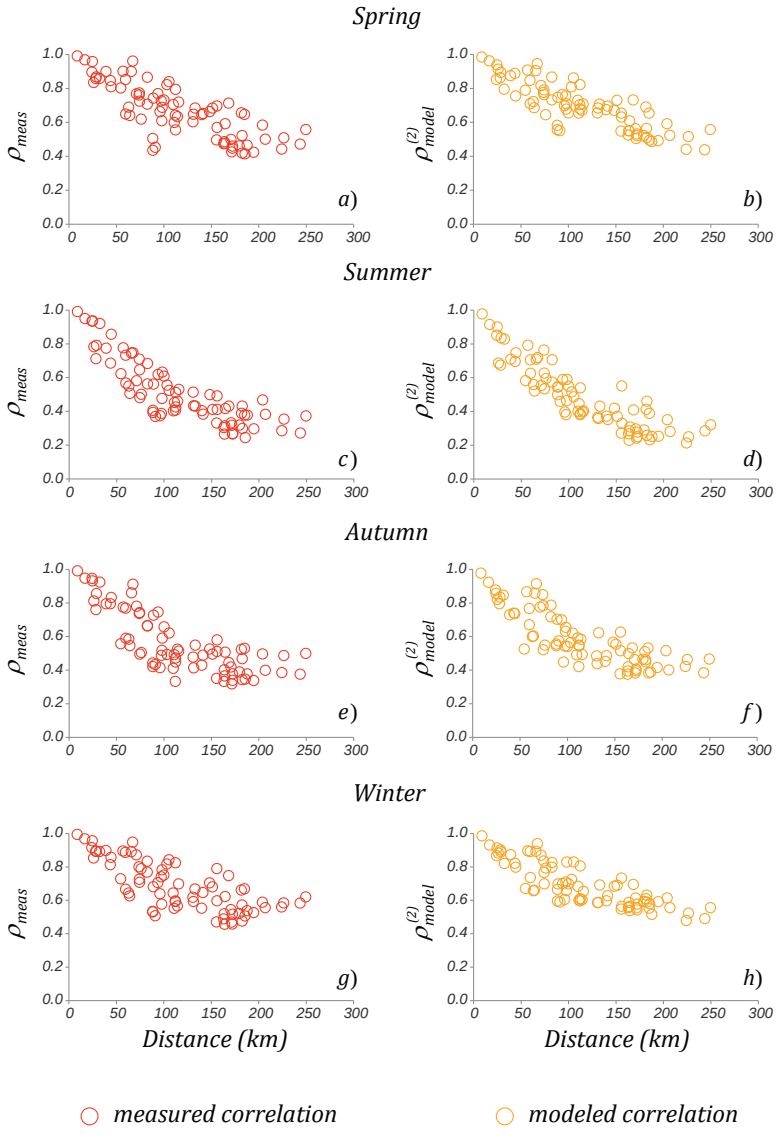


Figure 12: The seasonal streamflow correlation decreases with increasing catchment distance (distance measured from the centroids of the relevant basins). The Figure shows the observed streamflow correlation and the analytical streamflow correlation obtained using equation (30) with parameters estimated from discharge as a function of the inter-catchment distance.

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, inter-catchment differences in terms of land

cover (vegetation) and water retention capacity is possibly enhanced by higher evapotranspiration rates, increasing the inter-catchment 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 and 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 13, where equation (30) is employed. A general shift towards more correlated streamflows is observed in winter and spring, whereas a general loss of correlation, and a visible increase in the inter-catchment variability is observed in autumn and summer. Performances are satisfactory across all the four season (see also boxplots in Figure 13), 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.

### 3.5.2 *Analysis of the impact of inter-catchment 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 Figures 14 represent the frequency distribution of the factors $F_{\lambda\alpha}$ and $F_k$ in equations (29) and (30) for the 312 pairs of case studies. Given the multiplicative nature of these equations, values of $F$

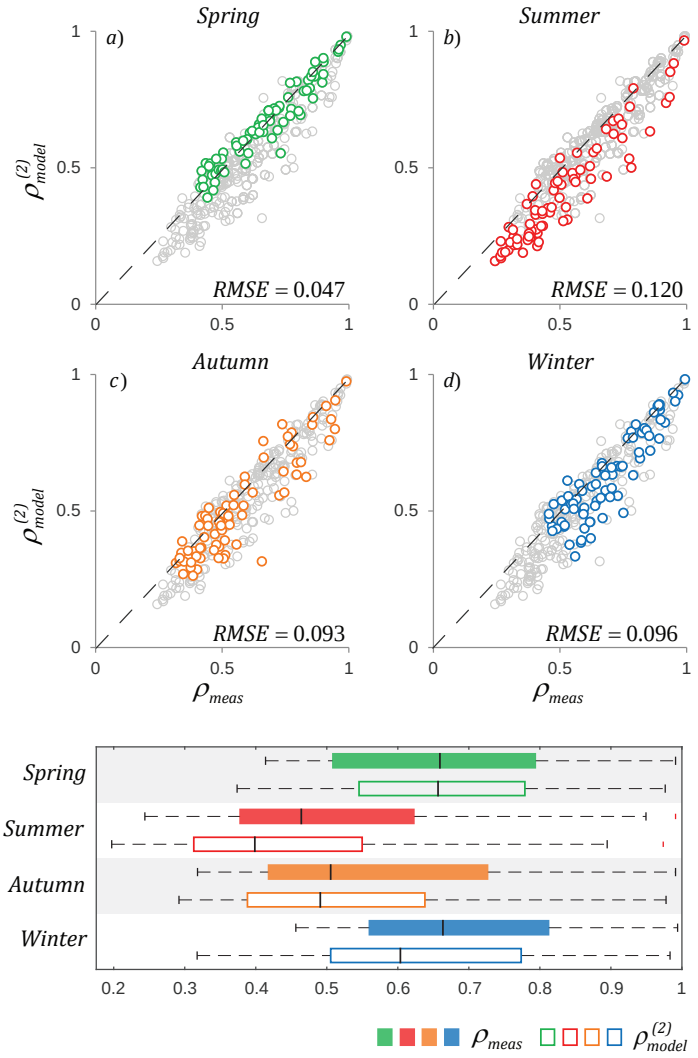


Figure 13: The scatterplots show the seasonal shift of streamflow correlation throughout the year. Summer and autumn display generally lower and more heterogeneous streamflow correlation, whereas streamflow dynamics are generally more correlated during spring and winter. Here, equation (30) with streamflow-estimated 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.

approaching 1 indicate a limited impact on the correlation of daily flows. Figure 14 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

streamflow correlation can be induced by enhanced differences between the recession rates in the two catchments ($\frac{k_1}{k_2} > 10$).

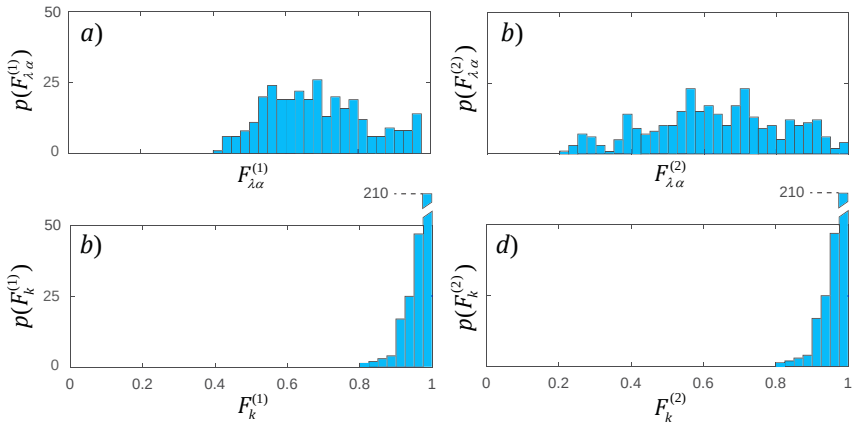


Figure 14: Frequency distributions of $F^{(1)}_{\lambda\alpha}$ and $F^{(1)}_{k}$ (equation (29)) and $F^{(2)}_{\lambda\alpha}$ and $F^{(2)}_{k}$ (equation (30)) for the 312 estimates of seasonal streamflow correlations. Values of $F_*$ close to 1 mean that the hydrological process represented by $*$ 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 (31) are similar to those of equation (30) and hence they are not shown here.

Figure 15 investigates how the heterogeneity of catchment-scale hydrological drivers (left panels), propagates to the streamflow correlation through the corresponding factors $F_\lambda$, $F_\alpha$ and $F_k$ (right panels). The analysis is carried out using equation (29) whose parameters are evaluated as detailed in section 3.4.2. 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 $\lambda_t$, $\alpha^{12}$ 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 panels of Figure 15. Additionally, the frequency distributions of $r_\alpha$ and $\frac{\lambda_{12}}{\lambda_m}$ ($\lambda_m = \min[\lambda_{1t}, \lambda_{12t}]$) are shown. In the left panels of Fig. 15, 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 15e shows that recession rates span a wide range of inter-catchment heterogeneity. However, because of the reduced sensitivity of $\rho_{model}$ on $V(k)$, the impact of inter-catchment variability of recession rates on the correlation of daily flows is quite limited (Figure 15f). On the contrary, the frequencies of effective rain-

fall 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 15a). However, the impact of the heterogeneity of rainfall frequencies on the streamflow correlation is significant (Figure 15b) because of the high sensitivity of $\rho_{model}$ 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 15c), 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.

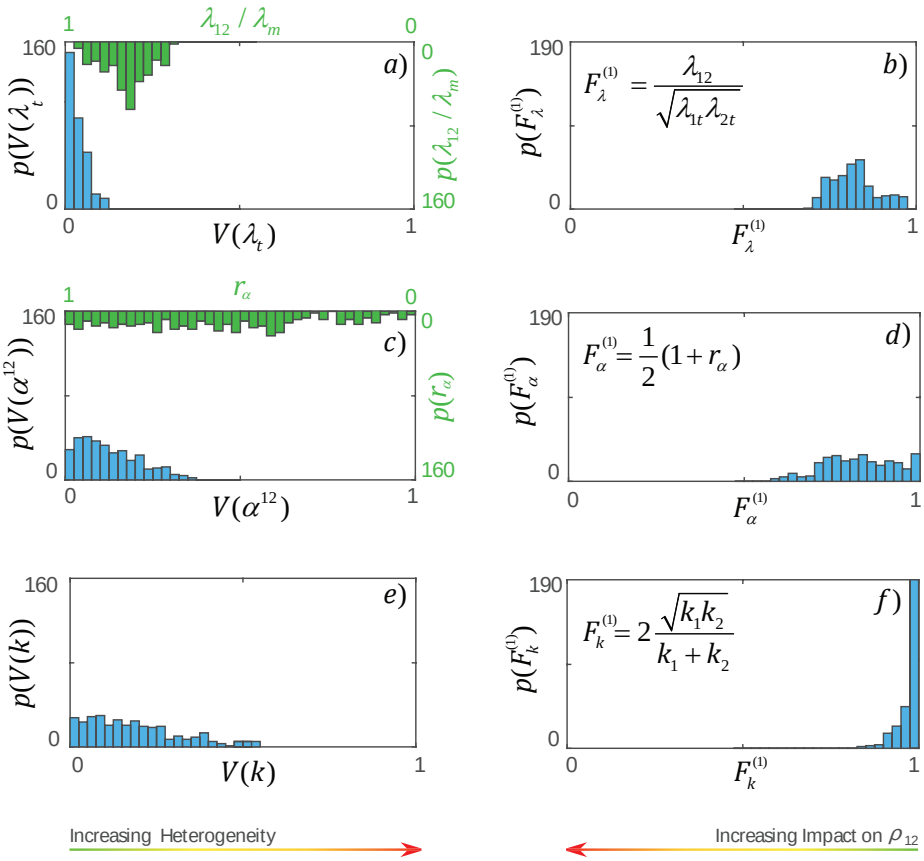


Figure 15: Frequency distributions of the variability index $V(*) := \frac{|*_1 - *_2|}{*_1 + *_2}$ ($* = \lambda_t$, $\alpha^{12}$, k, and 1, 2 identify the relevant catchments) and of the factors $F_\lambda^{(1)}$, $F_\alpha^{(1)}$ and $F_k^{(1)}$ for the 312 couples of estimated seasonal streamflow correlations. Equation (29) 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 inter-catchment heterogeneity in the corresponding hydrological descriptor $*$.

### 3.5.3 *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 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 and Botter, 2012; Lazzaro et al., 2013; Perona et al., 2013; Gorla and Perona, 2013*].

Streamflow PDF can be directly estimated from empirical frequency distributions [*Vogel and Fennessey, 1995; Castellarin et al., 2004, 2007*]. Unfortunately, streamflow gauging station 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 individuate 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 and Vogel, 2010*]. Therefore, streamflow correlation could be efficiently used to identify river sites having analogous flow regimes.

Equation (29) with rainfall-estimated parameters (see section 3.4.2) 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 16 shows the comparison between the normalized seasonal 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 ($q >> \langle q \rangle$).

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,

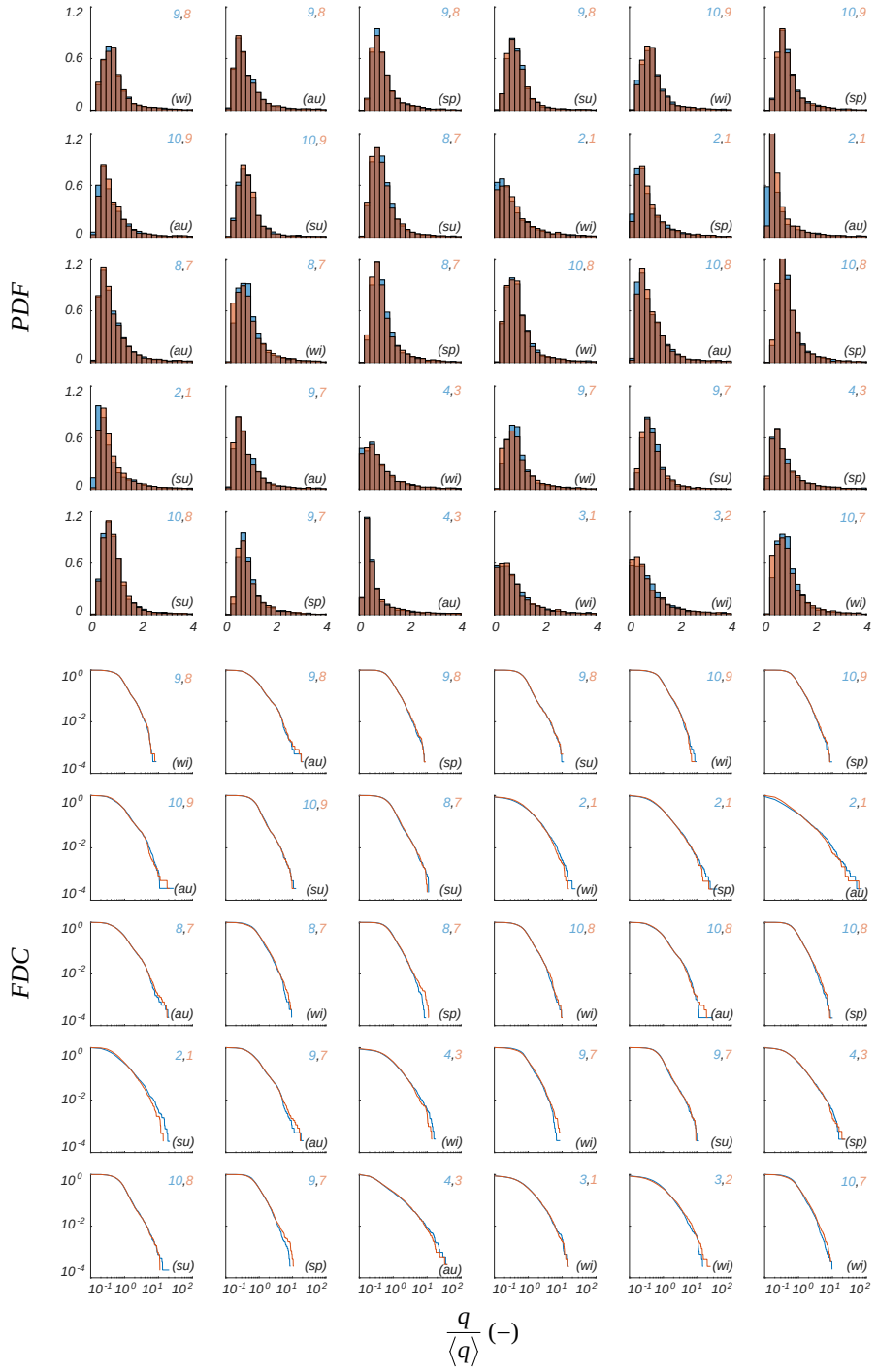


Figure 16: Comparison between the normalized seasonal streamflow PDF and FDC observed at all catchment outlets expected to have high streamflow correlation ($\rho^{(1)}_{model} > 0.9$). The couple of catchments considered in each plot is displayed (the numbers refer to catchments as in Tab. 4 and Fig. 7) as well as the corresponding season (in brackets). Highly correlated catchments have very similar flow statistics.

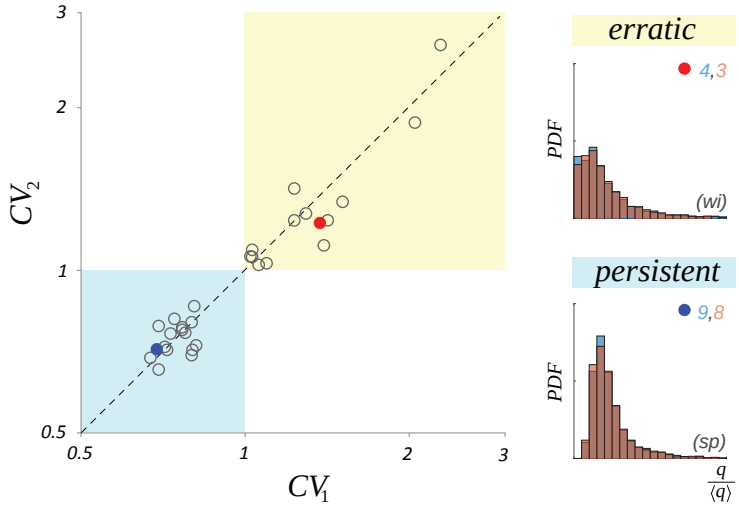


Figure 17: Comparison between the seasonal streamflow coefficient of variation (CV) between the pairs of outlets with $\rho^{(1)}_{model} > 0.9$. The framework is able to properly distinguish between erratic (CV $> 1$) and persistent (CV $< 1$) flow regimes (see text).

requiring only daily meteorological data readily available in most settings. The limited discrepancies between each couple of PDFs (FDCs) in Figure 16 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 17 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 absence of discharge data.

## 3.6 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 $75,000\,\mathrm{km}^2$ 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 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:

- 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.

- Equation (30) 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 (31) with parameters estimated from discharge records provided performances similar to those of equation (30).

- Equation (29) 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.

- Correlation exhibits significant seasonal variability. On average, higher correlations are observed in winter and spring, whereas lower correlations and higher inter-catchment 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.