Physically consistent conceptual rainfall–runoff model for urbanized catchments

Mohamed Saadi, Ludovic Oudin, Pierre Ribstein

Physically consistent conceptual rainfall–runoff model for urbanized catchments

Mohamed Saadi*, Ludovic Oudin, Pierre Ribstein

Sorbonne Université, CNRS, EPHE, UMR METIS, F-75000, Paris, France.

Published in Journal of Hydrology, doi: 10.1016/j.jhydrol.2021.126394.

*Corresponding author: mohamed.saadi@sorbonne-universite.fr

Abstract

Hydrological models should be tested and evaluated for a wide variety of levels of urbanization before they are used to predict the impact of urbanization on catchment behavior. In this study, we illustrate a top–down approach of modifying step by step an hourly conceptual model structure (GR4H) to account for urbanization features. Modifying the original model structure included accounting explicitly for runoff from impervious surfaces by bypassing the soil moisture reservoir and varying the partitioning between quick flow and slow flow. These adaptations were chosen based on the reported specificities of urbanized catchments, namely, decreasing infiltration, increasing runoff, and fast runoff dynamics. Using a split-sample test, the relevance of each modification with regard to the reproduction of catchment response (i.e., observed streamflow) was assessed for a large sample of 273 urbanized catchments, located in France and the United States, for which mean total impervious area (TIA) varied between 0.05 and 0.59. Six continuous and three event-based criteria were used, and two statistical tests were applied to assess the significance of improvements. Results showed the following: **(i)** Tested modifications improved the ability of the model to reproduce the catchment response, especially high flows and observed streamflow amid dry conditions. **(ii)** Event-based evaluation using more than 45,000 events showed an improvement in predicting the event peak flow and event runoff volume, whereas no significant improvements were obtained in predicting the timing of peak flow. **(iii)** Newly added parameters were moderately to highly correlated with TIA, especially the calibrated proportion of impervious surfaces, which is promising as a hydrological validation of estimated urbanization measures from land cover. The tested modifications improved both the representation of urbanization processes and the reproduction of the observed streamflow, yielding a simple and credible model for predicting the impact of future urbanization scenarios on catchment response.

Keywords: Urban hydrological model, model development, top–down approach, land-use change, model evaluation, GR4H

1 Introduction

1.1 Hydrological models as valuable tools for assessment of urbanization impact

There is strong evidence that urbanization modifies the hydrological behavior of catchments (Braud *et al.*, 2013; Fletcher, Andrieu, and Hamel, 2013; Leopold, 1968; McGrane, 2016; Miller and Hess, 2017). Nonetheless, predicting and quantifying the impact of urbanization on the rainfall–runoff relationship at the catchment scale is still a challenge (Oudin *et al.*, 2018; Redfern *et al.*, 2016). To this end, two types of approaches are generally applied (Braud *et al.*, 2013; Salavati *et al.*, 2016): a statistical approach and a modeling approach. The statistical approach seeks to either **(i)** identify temporal trends in the behavior of a catchment provided that long periods of hydroclimatic data are available across the urbanization period over the catchment area (Haase, 2009), or **(ii)** compare the behavior of an urbanized catchment with a non-urbanized one that has similar climatic and geomorphological characteristics (e.g., slope, elevation, lithology), as in paired-catchment experiments (Bonneau *et al.*, 2018; Prosdocimi, Kjeldsen, and Miller, 2015). These approaches are useful to detect and quantify past changes, but they do not provide physically sound links between hydrological processes and urbanization, which are necessary to reliably predict how the catchment behavior would be altered under future urbanization scenarios. In this respect, the modeling approach is advantageous because it backs the statistical method by providing a synthesis of the catchment behavior via the model parameters. This helps not only to detect change in catchment behavior (Saadi, Oudin, and Ribstein, 2020b; Pathiraja *et al.*, 2018), but also to create hydrological scenarios that correspond to urbanization scenarios (Niehoff, Fritsch, and Bronstert, 2002; McIntyre *et al.*, 2014; Sanzana *et al.*, 2019; De Niel *et al.*, 2020).

1.2 Overview of hydrological models for urbanized catchments

There is a spectrum of modeling tools along which a compromise is made between detailed representation of the spatial variability of hydrological processes and model simplicity (Hrachowitz and Clark, 2017; McIntyre *et al.*, 2014; Salvadore, Bronders, and Batelaan, 2015). Distributed models that use small-scale equations to represent the main hydrological processes are applied to account for the highly heterogeneous nature of hydrological processes in urban areas, which is accompanied by rapid dynamics of runoff generation on impervious surfaces (Cristiano, Veldhuis, and Giesen, 2017; Ogden *et al.*, 2011; Salvadore, Bronders, and Batelaan, 2015). Their application is also advocated because they explicitly represent catchment properties, which enables a direct assessment of the impact of urbanization by changing model parameters and scaling up the impact (Bronstert, Niehoff, and Bürger, 2002; Beven, 2002). This generally leads to heavily parametrized model structures (e.g., Cuo *et al.*, 2008; Jankowfsky *et al.*, 2014; Jia *et al.*, 2001; Sanzana *et al.*, 2019; Stavropulos-Laffaille *et al.*, 2018), which impedes testing their ability to reproduce the rainfall–runoff relationship for many catchments with a mix of rural and urban areas. Also, constraining these models requires a large volume of data, available only for a handful of monitored catchments (Rodriguez, Andrieu, and Creutin, 2003; Petrucci and Bonhomme, 2014).

At the other side of the spectrum, lumped, conceptual models that rely on relatively simple parametrization of water fluxes to describe the catchment-scale manifestation of small-scale heterogeneities are more easy to implement and less data-demanding. Thus, they offer the possibility of testing their robustness under different climate and land-cover situations, as has been shown by numerous model testing experiments using large samples of catchments (Gupta *et al.*, 2014). Perrin, Michel, and Andréassian (2001) compared 19 lumped daily models across 429 catchments located in France and the United States to discuss the issue of model complexity. Le Moine *et al.* (2007) tested different model modifications to account for daily inter-catchment groundwater flows on 1040 catchments located in France. At the hourly time step, Esse *et al.* (2013) compared the fixed and the flexible conceptual modeling approaches using 237 catchments located in France. Recently, Ficchì, Perrin, and Andréassian (2019) improved the consistency of the fluxes of a conceptual model across multiple sub-daily time steps using a set of 240 French catchments. These and many other studies focusing on model development, regionalization, or evaluation, as reviewed by Gupta *et al.* (2014), were mostly concerned with non-urbanized catchments, a fact that is mirrored by the existing large samples of catchment-scale hydroclimatic data that did not necessarily focus on (or even excluded) highly urbanized cases (Addor *et al.*, 2020).

Despite their relative simplicity, most of existing applications of conceptual models to urbanized catchments were limited to only few places at each time (Salvadore, Bronders, and Batelaan, 2015), highlighting the need for intensive testing of conceptual models using many urbanized catchments. Huang *et al.* (2008) quantified the impact of increasing imperviousness in the Wu-Tu catchment, Taiwan, on peak flow recurrence using a combination of Nash model and the Curve Number (CN) method. Dotto *et al.* (2011) conducted a sensitivity analysis of MUSIC and KAREN models on five urban Australian catchments with different imperviousness levels. Recently, De Niel *et al.* (2020) developed a methodology based on the NAM model to incorporate and project the impact of rapid urbanization on the hydrological behavior of two Belgian catchments, and Fidal and Kjeldsen (2020a) improved the conceptual model URMOD to account for soil moisture across 28 urbanized catchments in United Kingdom. Nonetheless, direct projection of the impact of urbanization on catchment behavior using conceptual tools is still undermined by the lack of explicit links between their structures and landscape properties. One could cite the CN method as an exception, but its use for continuous applications is impeded by inconsistencies in its formulation (Michel, Andréassian, and Perrin, 2005), and its ability to predict the impact of land-use changes (including urbanization) on catchment response is generally unverified (Ogden *et al.*, 2017).

A promising way of developing robustly tested models while at the same time enabling an explicit link between model structure and urbanization features lies in modifying already existing conceptual models to incorporate some physical properties of catchments (Euser *et al.*, 2015; Gharari *et al.*, 2014; Hrachowitz *et al.*, 2014; Kirchner, 2006). This can be guided by learning about the behavioral specificities of urbanized catchments using hydrological signatures (Gupta, Wagener, and Liu, 2008; McMillan *et al.*, 2011; Saadi, Oudin, and Ribstein, 2020b). An example was shown by Kjeldsen, Miller, and Packman (2013), who modified a non-urban model structure to account for urbanization features. They explicitly linked the model parameters (up to four) to the proportion of urban cover in the catchment, but their test concerned only seven catchments located in the United Kingdom. Another study by Hamel and Fletcher (2014) illustrated an improvement of the representation of low-flow components (interflow and groundwater flow) by gradually adding reservoirs to distinguish the contributions of the different parts of an urbanized catchment (including the impervious part of stream area and riparian zone). Their model modifications helped them test different stormwater management strategies to analyze their role in mitigating the impact of catchment imperviousness on baseflow. They defined their model modifications by analyzing the hydrological signatures of the McMahons Creek catchment in Australia. Both studies illustrate a top–down approach of improving the representation of urban hydrological processes in conceptual models, although their application was constrained to small sets of catchments. Thus, an attempt is required to balance the simplicity of conceptual models with robust model assessment using large samples of catchments, in order to detect their weaknesses and increase their credibility (Andréassian *et al.*, 2009; Gupta *et al.*, 2014; Klemeš, 1986), especially in the case of catchments with changing landscape.

1.3 Research gap, novelty and scope of the paper

Our research motivation stems from the fact that currently available models for urbanized catchments lack intensive testing on large number of cases. Among the 43 model applications in urbanized catchments reported in the exhaustive literature review by Salvadore, Bronders, and Batelaan (2015), none used more than 6 catchments to develop/test their modeling tool. We argue that using a large sample of urbanized catchments is a necessity to **(i)** advance our general understanding of how catchments with a mix of non-urban (or pervious) and urban (or impervious) covers behave hydrologically, by building on what we already know from non-urbanized ones, and thus **(ii)** make a credible extrapolation of catchment behavior under future urban planning schemes. Simple conceptual models developed for non-urbanized catchments have been proved to be flexible enough to adapt their model parameters from rural to urbanized contexts since the hydrological processes are not that different (Fletcher, Andrieu, and Hamel, 2013; Redfern *et al.*, 2016; Saadi, Oudin, and Ribstein, 2020b). Adapting their structures to urbanized environments should fulfill the following requirements: **(i)** an equal or improved ability (relative to the original conceptual model) to simulate the response of urbanized catchments (Fidal and Kjeldsen, 2020b), and **(ii)** more explicit links between the model parameters/structure and the urban characteristics of the catchments in order to enhance the physical consistency of the conceptual models with respect to landscape specificities (Hrachowitz *et al.*, 2014; Gharari *et al.*, 2014).

To fill this gap, we used a large sample of 273 urbanized catchments located in France and the United States, for which the mean total impervious area (TIA) ranged between 0.05 and 0.59. We conducted a step-by-step modification of the hourly non-urban GR4H model (Ficchì, Perrin, and Andréassian, 2019). Each modification was evaluated using a set of six continuous and three event-based evaluation criteria, and two statistical tests were applied to evaluate the statistical significance of each improvement with regard to the original non-urban model structure (Fidal and Kjeldsen, 2020b). By the modifications, we aimed at improving the physical consistency of the original conceptual structure by **(i)** testing modifications that were in agreement with the behavioral specificities of urbanized catchments, and **(ii)** comparing the added model parameters with observable urban characteristics.

This paper is organized as follows: In Section 2.1, we present the catchment set. Section 2.2 details the tested modifications of model structure to account for urbanization, and Section 2.3 describes the framework of their calibration and evaluation. Results are presented in Section 3, followed by a discussion and some perspectives in Section 4.

2 Dataset and methods

2.1 Catchment sample

We selected a large sample of 273 urbanized catchments, located in the United States (US) and France, based on four criteria (Saadi, Oudin, and Ribstein, 2019):

1. Availability of at least 8 years of hourly streamflow, precipitation, and daily temperature between 1997 and 2017. We considered that a minimum of 4 years are required for model calibration (Perrin *et al.*, 2007; Merz, Parajka, and Blöschl, 2009).

2. Limited snow influence, as snow melting was not addressed in the tested model.

3. Limited impact of large artificial reservoirs, which required streamflow naturalization (Terrier *et al.*, 2021).

4. A minimum mean total impervious area (TIA) of 0.05. This choice aimed at including the lowest level of catchment imperviousness above which the impact of urbanization on catchment behavior is often considered significant (Saadi, Oudin, and Ribstein, 2020a; Salvadore, Bronders, and Batelaan, 2015; Booth and Jackson, 1997; Arnold and Gibbons, 1996). Moreover, it offered a wide range of catchment imperviousness for a more robust model testing.

For the catchments located in the United States, we obtained the hourly precipitation depths via the `geoknife` R package (Read *et al.*, 2015) from the Stage IV dataset produced by the National Centers for Environmental Prediction (NCEP; Lin and Mitchell, 2005). We processed the hourly streamflow depths from the instantaneous hydrographs, which we extracted using the `dataRetrieval` R package (Cicco *et al.*, 2018) from the dataset of gauges maintained by the US Geological Survey (Falcone, 2011). We used the Daymet dataset to compute daily temperature time series (Thornton *et al.*, 2016). Land cover and imperviousness were characterized using the National Land Cover Database (NLCD) available for the years 2001, 2006, 2011, and 2016 (Wickham *et al.*, 2014; Homer *et al.*, 2007; Homer *et al.*, 2015). The location of the 205 US catchments is shown in Figure 1a.

For the catchments located in France, we processed the hourly precipitation depths from the COMEPHORE product of Météo France (Tabary *et al.*, 2012). We extracted the hourly streamflow depths from the Banque HYDRO dataset (Leleu *et al.*, 2014). Daily temperature was provided by the SAFRAN product of Météo France (Vidal *et al.*, 2010). Catchment imperviousness was characterized using the Imperviousness Density layers of the Copernicus Land Monitoring Service (CLMS), available for the years 2006, 2009, 2012, and 2015 (e.g., Congedo *et al.*, 2016). The location of the 68 French catchments is shown in Figure 1b.

In addition to precipitation, potential evapotranspiration was required as climatic forcing. It was estimated at daily time steps using a temperature-based formula (Oudin *et al.*, 2005). Then, an hourly disaggregation using the hourly extraterrestrial radiation helped estimate the hourly values (Allen *et al.*, 1998).

NLCD products mapped imperviousness over each 30-m pixel in the United States based on the land-cover type of the pixel. They assigned an imperviousness value of 0% to non-urban land cover types, whereas urban land-cover types were assigned an imperviousness value depending on their development intensity (Homer *et al.*, 2007; Homer *et al.*, 2004): **(1)** less than 20% for *Developed, Open Space*; **(2)** between 21% and 50% for *Developed, Low Intensity*; **(3)** between 51% and 80% for *Developed, Medium Intensity*; and **(4)** between 81% and 100% for *Developed, High Intensity*. For France, the CLMS high-resolution imperviousness products used a multi-linear regression model to predict imperviousness values over each 20-m pixel using biophysical variables, namely the Normalized Difference Vegetation Index (NDVI). This model assigned values between 1% and 100% to pixels that were classified as "built-up", whereas "non-built up" areas were assigned an imperviousness value of 0% (Langanke *et al.*, 2016). For each catchment and for each year of land cover data, we extracted the imperviousness values from the NLCD and CLMS imperviousness products using the

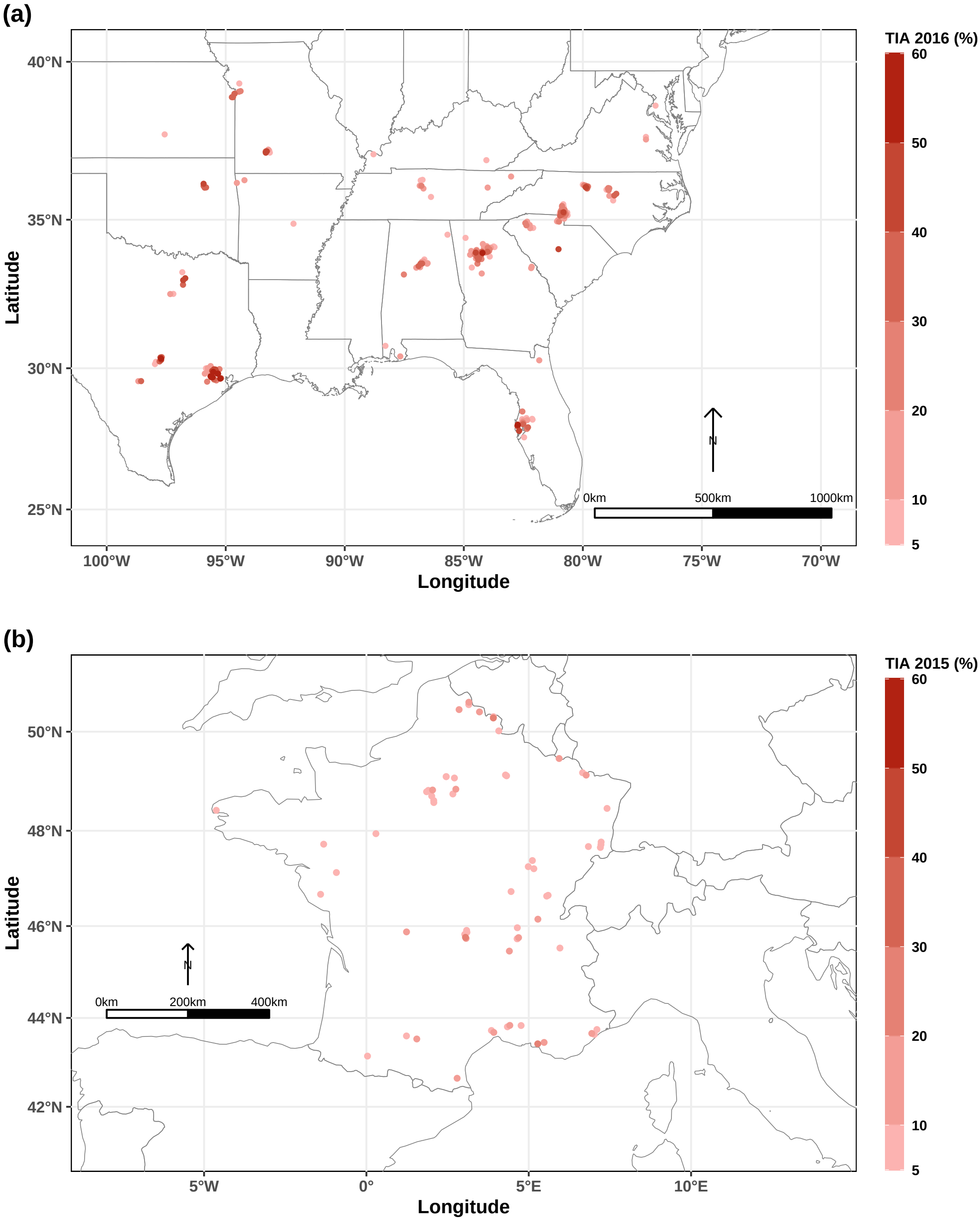

Figure 1. **(a)** Location of the 205 US catchments and **(b)** location of the 68 French catchments. A dot represents the location of the catchment centroid. Colors represent the mean total impervious area (TIA) of the catchment **(a)** for the year 2016, extracted from the National Land Cover Database, and **(b)** for the year 2015, extracted from the Imperviousness Density layers of the Copernicus Land Monitoring Service.

Table 1. Summary of catchment characteristics.

Catchment attribute	Min	Median	Max
Catchment area (km^2)	1.1	60	2100
Length of period of records (years)	8	16	16
Mean annual precipitation P (mm/year)	510	1170	1660
Mean annual potential evapotranspiration E (mm/year)	620	1030	1400
Humidity index P/E (–)	0.6	1.1	2.2
Mean TIA per period of records (–)	0.05	0.16	0.59

catchment polygon, and then we considered the arithmetic mean of the imperviousness values as catchment-scale TIA. For the missing years, we used a linear interpolation to complete the catchment-scale TIA time series for the whole period of hydroclimatic records.

For event-based assessment, we extracted a set of events from the precipitation and streamflow time series using an empirical method (Ficchì, Perrin, and Andréassian, 2016; Lobligeois *et al.*, 2014; Saadi, Oudin, and Ribstein, 2020a). This method was applied to the direct flow time series to distinguish the events originating from surface runoff. Direct flow was determined by applying a numerical filter to extract the baseflow (Mei and Anagnostou, 2015; Collischonn and Fan, 2013; Eckhardt, 2005; Blume, Zehe, and Bronstert, 2007). This method yielded a total of 45,025 events, with a median number of events per catchment of approximately 140 events (interquartile range: 83–239).

Most of the US catchments used in this study are characterized by a temperate humid climate (Cfa type in the Geiger–Köppen classification, see Beck *et al.*, 2018), while the French catchments are characterized by temperate oceanic (Cfb) and Mediterranean climates (Csa). The relatively wide range of humidity indices, as summarized in Table 1, reflects the climatic diversity of the dataset. Also, the 273 catchments contain a variety of urbanization levels, manifested by a wide range of mean TIA.

2.2 Model modifications

2.2.1 Original model structure

Urbanized catchments are characterized by fast response dynamics due to surface sealing and direct conveyance through artificial drainage systems (Salvadore, Bronders, and Batelaan, 2015; Rodriguez, Andrieu, and Creutin, 2003; Fletcher, Andrieu, and Hamel, 2013). Hence, the choice of hourly to sub-hourly time steps is necessary in order to accurately describe the catchment response. However, the paucity of hydroclimatic datasets at fine time steps imposes a maximum time resolution. Considering both constraints, we chose to test our modifications on a model running at the hourly time step.

The GR4H model (Ficchì, Perrin, and Andréassian, 2019) was selected as the starting point for model development, but the methodology is applicable to any other conceptual model with similar complexity. GR4H was not developed specifically for urbanized catchments, but it has been widely tested in large international sets of catchments and in model intercomparison studies (Boer-Euser *et al.*, 2017; Le Moine *et al.*, 2007; Esse *et al.*, 2013). In addition, it proved to be comparatively flexible when adapting its calibrated

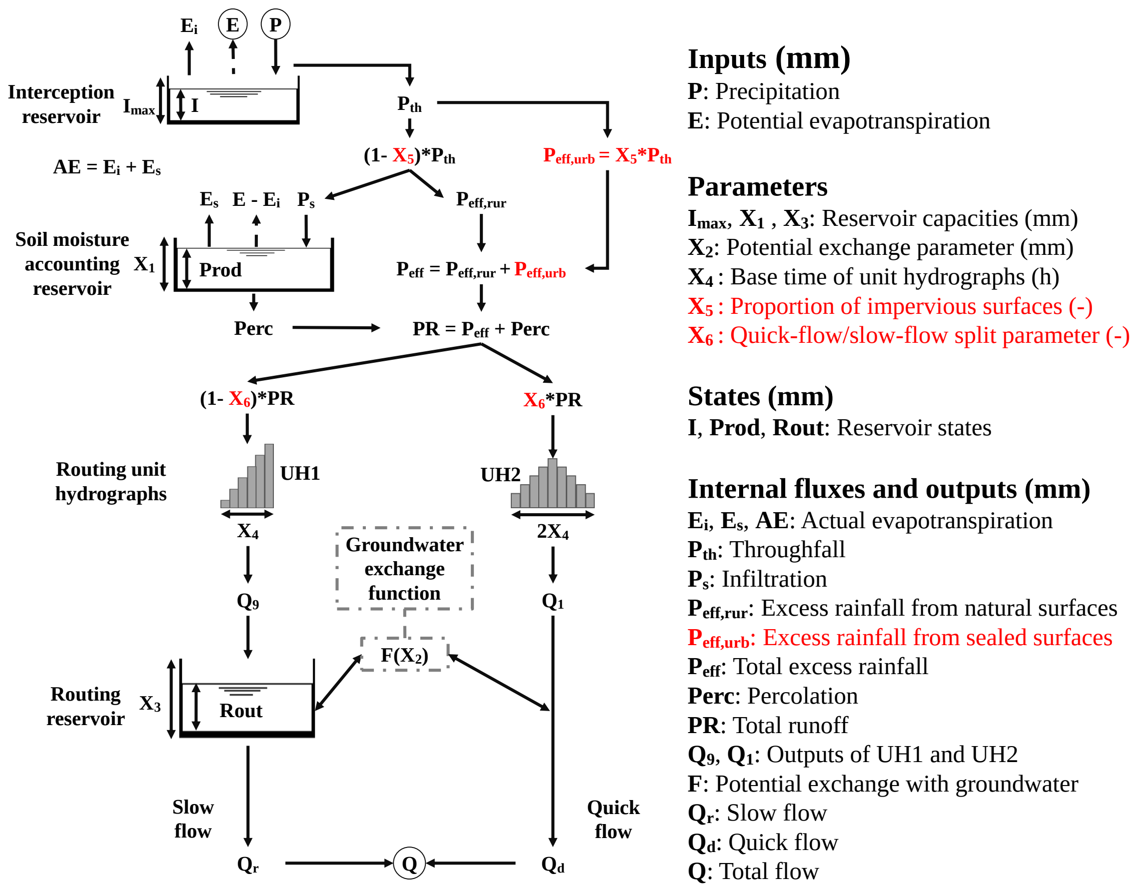

Figure 2. Description of GR4H model structure, parameters, states, and internal fluxes. In red, we highlight the modifications that we carried out to account for urbanization features. In the original model structure, X_5 is set at 0 (i.e., $P_{eff,urb}$ is null and $P_{eff,rur}$ is equal to P_{th}–P_s) and X_6 is fixed at 0.1.

parameters to urbanized catchments, leading to a relatively satisfactory reproduction of observed streamflow for a large spectrum of urbanized contexts (Saadi, Oudin, and Ribstein, 2020b). The structure of the model is shown in Figure 2. This structure was inherited from the daily version GR4J, for which a detailed description of equations is given by Perrin, Michel, and Andréassian (2003). The model was recently modified by Ficchì, Perrin, and Andréassian (2019) by adding an interception reservoir to guarantee a stable internal functioning across different time steps.

The interception reservoir is fed by the precipitation P and emptied to satisfy the potential evapotranspiration E. The throughfall P_{th} is generated when the interception reservoir is full. In this case, the soil moisture-accounting reservoir (also referred to as the "production reservoir") is fed by a depth P_s, function of P_{th}, $Prod$ and X_1, where $Prod$ is the updated state of the production reservoir and X_1 its capacity. The excess rainfall, i.e., P_{th}–P_s, is added to percolation $Perc$ from the production reservoir to constitute the net precipitation PR. The latter is split into $0.1 \cdot PR$ routed through the quick-flow branch (unit hydrograph UH2 with a base time $2 \cdot X_4$ hours), and $0.9 \cdot PR$ passed through the slow-flow branch (unit hydrograph UH1 with a base time X_4 hours, followed by a non-linear reservoir with updated state $Rout$ and capacity X_3). Before reaching the simulated total flow, a surface water–groundwater exchange algebraic depth $F(X_2)$, function of X_2 and the rate $\frac{Rout}{X_3}$, is added to the outflow of UH1 and the level of the routing reservoir ($Q_9 + Rout$) on the slow-flow branch, and to the outflow of UH2 (Q_1) on the quick-flow branch. $F(X_2)$ is positive (or negative) when water is imported from (or exported to) groundwater. Finally, the total flow is the sum of the quick flow Q_d and the slow flow Q_r, with the latter being the outflow of the routing reservoir, determined using $Rout$ and X_3.

This initial model structure contains four parameters that we estimated using numerical calibration: X_1 (mm) the capacity of the production store, X_2 (mm) the exchange function parameter, X_3 (mm) the capacity of the non-linear routing reservoir, and X_4 (h) the base time of the unit hydrographs. I_{max} (mm), the capacity of the interception reservoir, was estimated by reducing the differences between hourly and daily throughfall (Ficchì, Perrin, and Andréassian, 2019), where the latter is estimated at the daily time steps by comparing daily P and E (Perrin, Michel, and Andréassian, 2003). For the set of catchments used here, I_{max} varied between 1.5 and 4.0 mm, with a median value of 2.75 mm. The model states $Prod$ (the dynamic level of the production reservoir in mm) and $Rout$ (the dynamic level of the routing reservoir in mm) were initialized with $0.3 \cdot X_1$ and $0.5 \cdot X_3$, respectively, whereas the level of the interception reservoir was initialized with 0 mm. To limit the effect of these initializations on model performances, a warm-up period (one year) was used prior to each calibration/test period.

2.2.2 Accounting for the presence of impervious surfaces

Urbanization increases the proportion of impervious surfaces, which limits infiltration and increases runoff (Saadi, Oudin, and Ribstein, 2020a; Walsh *et al.*, 2005; Leopold, 1968; Miller and Hess, 2017; Zhou *et al.*, 2017). This is equivalent to establishing a sort of disconnection between catchment response (i.e., runoff) and the soil moisture state. To account for this effect, the throughfall P_{th} is split in two portions (see Figure 2): **(i)** a portion that bypasses the production reservoir and becomes excess rainfall immediately, $X_5 \cdot P_{th}$, which can be interpreted as the excess rainfall from impervious surfaces, and **(ii)** a portion that falls on the natural surfaces, $(1 - X_5) \cdot P_{th}$, and contributes to feeding the production reservoir. The parameter X_5 (–) is intended to represent the proportion of the impervious surfaces, which can be conceptually linked to TIA.

In the original structure, the ratio of instantaneous excess rainfall p_{eff} (mm/h) to instantaneous throughfall p_{th} (mm/h) is expressed as:

$$\frac{p_{eff}}{p_{th}} = \left(\frac{Prod}{X_1}\right)^2 \tag{2.1}$$

By including the excess rainfall on sealed surfaces, Equation 2.1 becomes:

$$\frac{p_{eff}}{p_{th}} = X_5 + (1 - X_5) \cdot \left(\frac{Prod}{X_1}\right)^2 \tag{2.2}$$

Under very dry conditions ($Prod \sim 0$), the proportion of throughfall that is converted into excess rainfall is as high as X_5 (Figure 3), i.e., as high as the mean catchment imperviousness. Note that with $X_5 = 1$, i.e., a completely impervious catchment, all throughfall bypasses the production reservoir and thus instantaneously reaches the transfer function. On the other hand, with $X_5 = 0$, we go back to the original formulation in Equation 2.1, i.e., the pre-urbanized era of the catchment, for which all excess rainfall is estimated based on the filling rate of the production reservoir $\frac{Prod}{X_1}$.

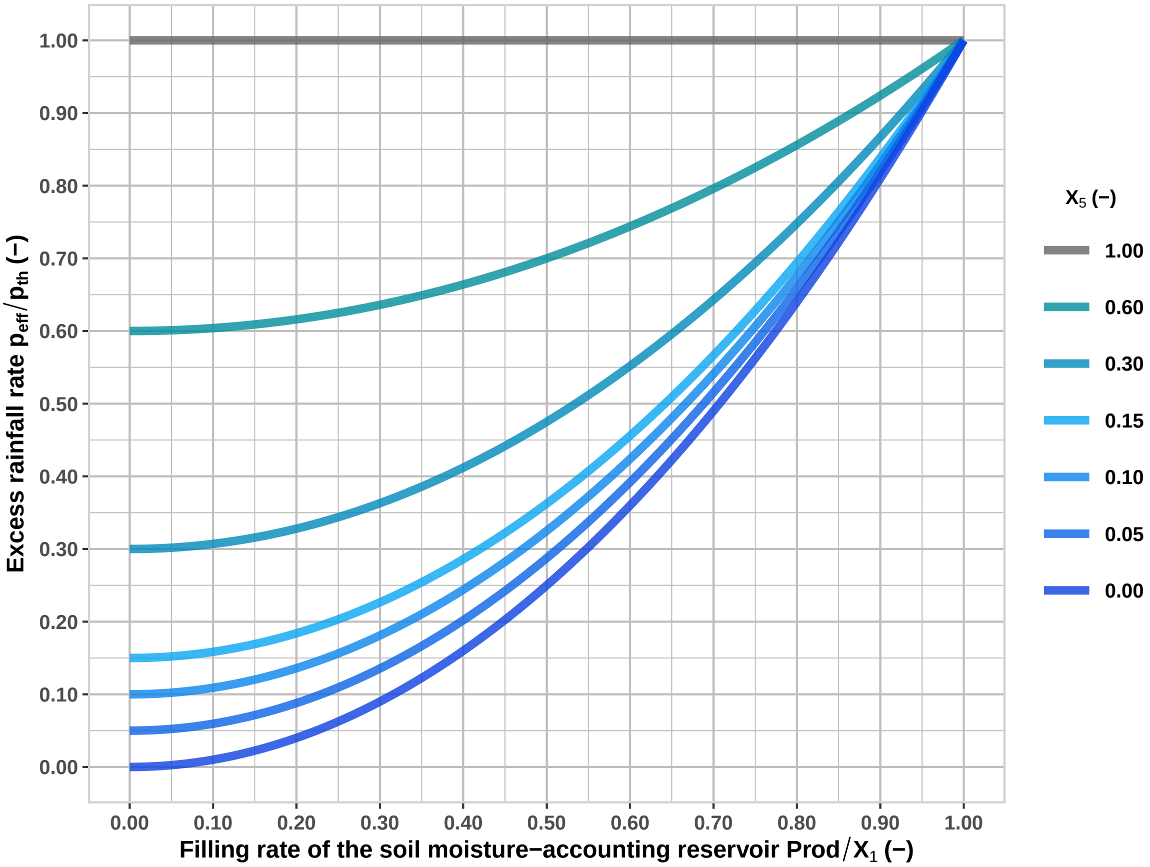

Figure 3. Illustration of the effect of different values of X_5, representing conceptually the proportion of impervious surfaces, on the relationship between the excess rainfall rate $\frac{p_{eff}}{p_{th}}$, and $\frac{Prod}{X_1}$, the filling rate of the soil moisture-accounting reservoir.

Equation 2.2 can be also viewed as:

$$\begin{aligned} p_{eff} &= X_5 \cdot p_{th} + (1 - X_5) \cdot p_{th} \cdot \left(\frac{Prod}{X_1}\right)^2 \\ &= p_{eff,urb} + p_{eff,rur} \end{aligned} \tag{2.3}$$

where $p_{eff,urb} = X_5 \cdot p_{th}$ is the excess rainfall generated in impervious/urbanized surfaces, and $p_{eff,rur} = (1 - X_5) \cdot p_{th} \cdot \left(\frac{Prod}{X_1}\right)^2$ is the excess rainfall coming from pervious/natural surfaces.

By integrating over a time step of 1 hour, the depth P_s (mm) that feeds the soil moisture reservoir becomes:

$$P_s = \frac{X_1 \cdot \left(1 - \left(\frac{Prod}{X_1}\right)^2\right) \cdot \tanh\left(\frac{(1-X_5) \cdot P_{th}}{X_1}\right)}{1 + \frac{Prod}{X_1} \cdot \tanh\left(\frac{(1-X_5) \cdot P_{th}}{X_1}\right)} \tag{2.4}$$

which is now dependent on X_5, in addition to X_1, P_{th} (p_{th} integrated over 1 hour), and $\frac{Prod}{X_1}$.

2.2.3 Accounting for the impact of urbanization on quick-flow/slow-flow partitioning

With increased urbanization, runoff dynamics are intensified (Miller and Hess, 2017; Saadi, Oudin, and Ribstein, 2020b; Diem, Hill, and Milligan, 2018), due to reduced surface roughness and artificial drainage systems. This results in higher amounts of net precipitation PR reaching the catchment outlet in a shorter span of time. In the original structure, the quick-flow/slow-flow split parameter X_6 was fixed at 0.1 (Figure 2), i.e., 10% and 90% of runoff give rise to quick flow and slow flow, respectively. Fixing this model parameter was chosen in the early model developments as its calibration on predominantly non-urbanized catchments did not result in significantly improved model performances (Perrin, Michel, and Andréassian, 2003). In urbanized catchments, this parameter would be expected to reach higher values than 0.1 since quick flow is generally quantitatively increased. Thus, we tested a model version for which this parameter is free and calibrated it for each catchment.

While X_5 can be conceptually linked to TIA, the product $X_5 \cdot X_6$, where X_6 is calibrated and X_5 is either fixed at TIA or calibrated, could be interpreted as the proportion of impervious surfaces that is directly connected to the catchment outlet. This is equivalent to average catchment-scale effective impervious area (EIA), considered to be a more relevant hydrological predictor than TIA (Alley and Veenhuis, 1983; Booth and Jackson, 1997; Lee and Heaney, 2003; Ebrahimian, Wilson, and Gulliver, 2016). Hence, we explored the relationship between the product $X_5 \cdot X_6$ and average TIA of calibration period by fitting an exponential function

$$X_5 \cdot X_6 = m \cdot TIA^n \tag{2.5}$$

where the parameters m (the scaling factor) and n (the exponent) were estimated by fitting linear least-squares regression models to log-transformed values of average TIA of calibration period and $X_5 \cdot X_6$. We compared the estimated values of m and n with previously proposed ones in the literature, such as the empirical equations of Sutherland (1995) whose scale factor and exponent parameters were given depending on the level of the connectedness of impervious surfaces to the drainage network (i.e., highly connected, average, somewhat disconnected, and extremely disconnected). Note that this analysis concerned only the models for which it was possible to both calibrate X_6 and change X_5 depending on each catchment (either through calibration or by fixing $X_5 = TIA$), i.e., models MU5H and MU6H (see Table 2).

2.2.4 Summary of tested models

Six model versions were compared across the 273 catchments, and their configurations are summarized in Table 2. Three possibilities were tested for the proportion of impervious surfaces X_5: **(1)** $X_5 = 0$ as in the original version, **(2)** X_5 is updated every year with *a priori* estimated TIA from land covers (i.e., the "actual" TIA), and **(3)** X_5 is set free and calibrated for each catchment. Moreover, two possibilities were tested for X_6, the quick-flow/slow-flow split parameter: **(1)** X_6 fixed at 0.1, as in the original version, and **(2)** X_6 calibrated on the whole calibration period. Note that calibrating X_6 for each year to accommodate the yearly change in TIA was not considered, as it would lead to an additional number of numerically calibrated parameters proportional to the number of years of calibration period.

Table 2. Summary of tested model modifications. The number of numerically calibrated parameters is given in each model name.

Model abbreviation	Urban production	Proportion of impervious surfaces X_5	Quick-flow/slow-flow split parameter X_6
GR4H (Original model structure)	***G**énie **R**ural à **4** paramètres au pas de temps **H**oraire* (Rural model with 4 parameters running at the Hourly time step)	Fixed, $X_5 = 0$	Fixed, $X_6 = 0.1$
MR5H	***M**odèle **R**ural à **5** paramètres au pas de temps **H**oraire* (Rural Model with 5 parameters running at the Hourly time step)	Fixed, $X_5 = 0$	Calibrated
MU4H	***M**odèle **U**rbain à **4** paramètres au pas de temps **H**oraire* (Urban Model with 4 parameters running at the Hourly time step)	Fixed, $X_5 = TIA$ estimated from land cover dataset	Fixed, $X_6 = 0.1$
MU5H	***M**odèle **U**rbain à **5** paramètres au pas de temps **H**oraire* (Urban Model with 5 parameters running at the Hourly time step)	Fixed, $X_5 = TIA$ estimated from land cover dataset	Calibrated
MUOpt5H	***M**odèle **U**rbain avec **Opt**imisation de la branche imperméable à **5** paramètres au pas de temps **H**oraire* (Urban Model including Optimized impervious branch with 5 parameters running at the Hourly time step)	Calibrated	Fixed, $X_6 = 0.1$
MU6H	***M**odèle **U**rbain à **6** paramètres au pas de temps **H**oraire* (Urbanized Model with 6 parameters running at the Hourly time step)	Calibrated	Calibrated

2.3 Evaluation of tested modifications

Retaining any of the model modifications should be warranted by a significant improvement of model performances on a large and diversified set of catchments (Andréassian *et al.*, 2009). The five model modifications (listed in Table 2) were assessed in comparison with the GR4H original model structure both on continuous and event bases, using a split-sample test (Klemeš, 1986). In this section, we detail **(i)** the model calibration procedure, **(ii)** the continuous and event-based assessment criteria, and **(iii)** the statistical tests that we applied to assess the significance of differences (improvements/deteriorations) in model performances.

2.3.1 Model calibration procedure

The period of records of each catchment was split into two subperiods P1 and P2 with equivalent lengths. For each subperiod and catchment, each model was calibrated by maximizing the objective function, chosen as the Kling–Gupta efficiency KGE (Gupta *et al.*, 2009) applied to the square root values of observed and simulated streamflow depths ($KGESQ$), in order to guarantee a uniform emphasis on all streamflow components (Santos, Thirel, and Perrin, 2018; Oudin *et al.*, 2006). $KGESQ$ is computed as:

$$KGESQ = 1 - \sqrt{(1-r)^2 + (1-\alpha)^2 + \left(1-\beta\right)^2} \qquad (2.6)$$

where r refers to the correlation of the observed and simulated time series, α the ratio of their standard deviations, and β the ratio of their means. $KGESQ$ ranges from $-\infty$ to 1, and the latter is the obtained value for an ideal simulation (i.e., for $r = \alpha = \beta = 1$). To this end, we broadly inspected the parameter hyperspace to look for a starting point, from which a gradient descent algorithm was followed to reach the optimal set of parameters that maximized $KGESQ$ (Coron *et al.*, 2017; Edijatno *et al.*, 1999).

2.3.2 Evaluation of model performaces using continuous and event-based assessment metrics

Model evaluation remains a difficult exercise as there is no perfect performance criterion (Krause, Boyle, and Bäse, 2005; Legates and McCabe, 1999), and focusing on only one evaluation criterion could mislead the modeler's choice (Schaefli and Gupta, 2007). For a rigorous assessment, we used two different metrics, $KGESQ$ and Nash-Sutcliffe Efficiency (NSE), and conducted an assessment during wet and dry conditions, in addition to an event-based assessment.

The obtained parameter set from calibration is tested on the alternative period (P1 if calibration on P2, and vice versa). In addition to $KGESQ$, continuous streamflow evaluation was conducted using NSE applied to observed and simulated streamflow depths, according to (Nash and Sutcliffe, 1970):

$$NSE = 1 - \frac{\sum_h \left(Q_h^{obs} - Q_h^{sim}\right)^2}{\sum_h \left(Q_h^{obs} - \overline{Q^{obs}}\right)^2} \qquad (2.7)$$

where Q_h^{sim}, Q_h^{obs} are simulated and observed discharges at hour h, respectively, and $\overline{Q^{obs}}$ is the mean observed discharge. NSE varies from $-\infty$ to 1, and the latter is the obtained value with identical simulated and observed streamflow time series.

Contrasts between urbanized and non-urbanized catchments are reported to be more significant during dry periods (Zhou *et al.*, 2017; Sillanpää and Koivusalo, 2015). Thus, $KGESQ$ and NSE were also computed with emphasis on wet ($KGESQ_{wet}$ and NSE_{wet}) and dry ($KGESQ_{dry}$ and NSE_{dry}) conditions of the test period. To do this, we multiplied the observed and simulated streamflow time series at each hour h by the weights $wwet_h$ and $wdry_h$ (Oudin *et al.*, 2006), computed as:

$$\begin{cases} wwet_h &= \frac{1}{1+\exp\left(-25\cdot\left(\frac{Prod_h}{X_1}-0.5\right)\right)} \\ wdry_h &= \frac{1}{1+\exp\left(+25\cdot\left(\frac{Prod_h}{X_1}-0.5\right)\right)} \end{cases} \quad (2.8)$$

with $Prod_h$ estimated at hour h using the interception and production reservoirs of GR4H (Figure 2), which varies between 0 and X_1. When $\frac{Prod_h}{X_1}$ approaches 1 (i.e., very wet conditions), $wwet_h \sim 1$ and $wdry_h \sim 0$, and *vice versa*. Note that $wwet_h$ and $wdry_h$ are always between 0 and 1. Different values of X_1 between 10 mm and 750 mm were tested to see which configuration dampens the observed streamflow best in summer (i.e., June to August) with $wwet_h$, and in winter (i.e., December to February) with $wdry_h$ for all the 273 catchments (not shown here). A value of $X_1 = 200$ mm yielded the best trade-off and was chosen to compute $wwet_h$ and $wdry_h$ at each hour h for every catchment. This weighting scheme helped to account for the climatic specificities of each catchment in the definition of wet/dry periods instead of imposing fixed winter/summer months for all the catchments.

We are much more interested in the performances of the tested models relative to the original structure. Hence, for each continuous criterion $Crit$, a relative index R_{Crit} was used to assess the reduction in error due to each modification relative to the original model structure. R_{Crit} is computed as (Lerat *et al.*, 2012):

$$R_{Crit}(Modif, Ref) = \frac{Crit(Modif)-Crit(Ref)}{2-Crit(Modif)-Crit(Ref)} = \frac{\text{Error}(Ref)-\text{Error}(Modif)}{\text{Error}(Ref)+\text{Error}(Modif)} \quad (2.9)$$

with $Crit \in \{KGESQ, NSE, KGESQ_{wet}, NSE_{wet}, KGESQ_{dry}, NSE_{dry}\}$, $Modif$ is the modified model, and Ref is the reference model, in our case GR4H. R_{Crit} varies between -1 and 1, with positive values indicating an improvement and negative values indicating a degradation obtained when using the model $Modif$ relative to the original model Ref. Note that $R_{Crit}(Ref, Ref) = 0$. A better interpretation of R_{Crit} can be obtained by computing the corresponding reduction in error:

$$\begin{cases} \text{Reduction in error} &= 1-\frac{\text{Error}(Modif)}{\text{Error}(Ref)} = \frac{2\cdot R_{Crit}(Modif,Ref)}{1+R_{Crit}(Modif,Ref)} \\ \text{Error} &= 1-Crit \end{cases} \quad (2.10)$$

In addition, the ability of models to reproduce the hydrographs of events belonging to the test subperiods was assessed using the error in estimating event peak flow e_{Qp}, the error in predicting the timing of peak flow e_{tp}, and the volumetric efficiency VE (Criss and Winston, 2008), defined as:

$$\begin{cases} e_{Qp} &= \frac{Q_p^{sim}-Q_p^{obs}}{Q_p^{obs}} \\ e_{tp} &= \frac{h_p^{obs}-h_p^{sim}}{h_p^{obs}-h_{st}^{obs}} \\ VE &= 1-\frac{\sum_{h_{st}^{obs}\le h\le h_{end}^{obs}} |Q_h^{obs}-Q_h^{sim}|}{\sum_{h_{st}^{obs}\le h\le h_{end}^{obs}} Q_h^{obs}} \end{cases} \quad (2.11)$$

with Q_p^{sim}, Q_p^{obs} the simulated and observed peak flows, and h_p^{sim}, h_p^{obs} the hours of their occurrence. h_{st}^{obs}, h_{end}^{obs} are the starting and ending hours of each event. The ideal value for e_{Qp} and e_{tp} is 0, whereas the ideal value of VE is 1. Positive e_{Qp} indicates an overestimation of the peak flow, whereas positive e_{tp} indicates that the predicted peak flow occurs before the observed peak flow.

2.3.3 Statistical significance of model improvements

To assess the statistical significance of the obtained improvements by each model modification M' (Table 2) relative to the original model structure GR4H, we used two tests to compare M' and GR4H regarding each continuous and event-based evaluation criterion. First, we applied the one-tailed Wilcoxon signed-rank test (Wilcoxon, 1945), assuming that the compared samples are paired since the same catchment sample was used to test both models. This test helped us to accept/reject each modification by examining whether the differences in median scores were statistically different from zero (i.e., significantly higher/lower than zero depending on each criterion). Second, we conducted a one-tailed binomial statistical test (Clopper and Pearson, 1934; Fidal and Kjeldsen, 2020b), where the null hypothesis is set such as GR4H performs equally to or better than M', in other words, the proportion of subperiods/events in which M' outperforms GR4H is equal to or less than 50%. Rejecting this hypothesis implies that the model modification leads to a statistically significant improvement for the majority of cases. The probability that the modified model version M' outperforms the original version GR4H on a number of subperiods or events is given by the binomial distribution. Outperformance of M' with respect to GR4H is achieved with a positive relative index R_{Crit} (Equation 2.9), decreased absolute values of e_{Qp}, decreased absolute values of e_{tp}, and increased VE. The number of trials, i.e., how many times M' and GR4H were compared, is 546 (the number of subperiods) in the case of continuous assessment and 45,025 (the number of events) in event-based assessment.

3 Results

3.1 Calibration performances and distributions of calibrated parameters

Calibrating more parameters resulted in improved calibration performances. Moreover, accounting for TIA resulted in better calibration performances for the majority of catchments, even with no additional calibrated parameters (MU4H vs. GR4H), as shown by the distributions of R_{KGESQ} in Figure 4. The median reduction of error ranged between 2% (R_{KGESQ} = 0.01) and 11% (R_{KGESQ} = 0.06), and improvements were observed for the majority of cases. Calibrating only X_5 yielded higher calibration performances than calibrating only X_6 (MUOpt5H vs. MR5H). However, calibrating X_6 while fixing $X_5 = TIA$ (i.e., MU5H) resulted in higher calibration performances relative to the original model GR4H.

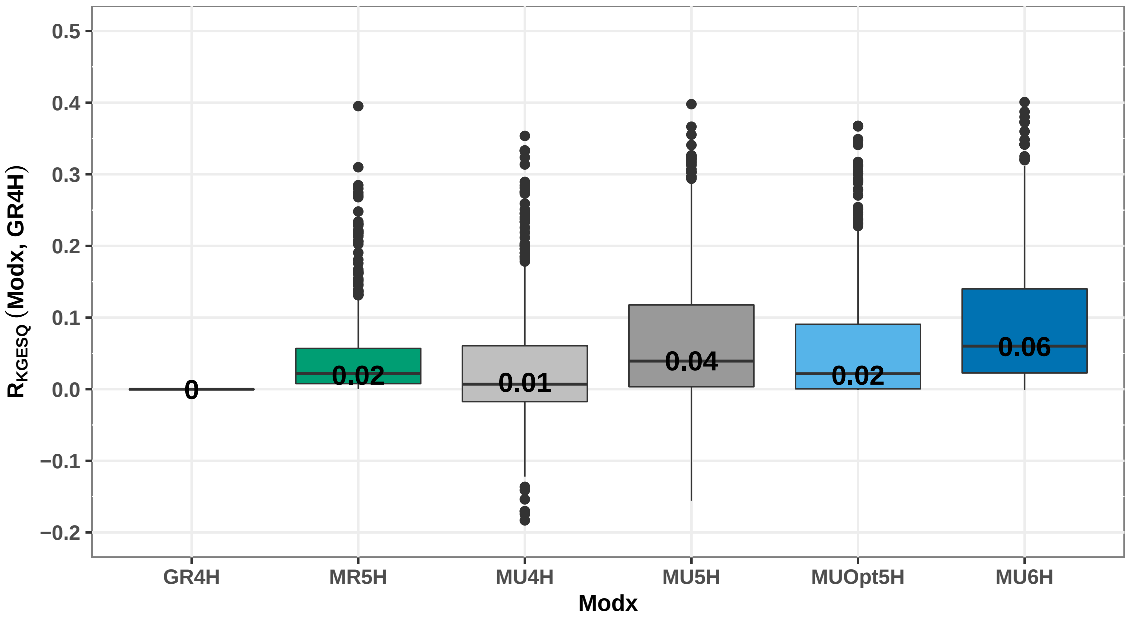

Figure 4. Distributions of improvements in calibration performances obtained for the six tested model structures, expressed in terms of relative R_{KGESQ}. Boxes are delimited by the first and third quartiles. Values indicate the median of each distribution.

Interestingly, calibrated values of X_5 in MUOpt5H and MU6H were highly correlated to observed values of TIA, as illustrated in Figure 5. The median value of mean TIA over the 546 subperiods was 0.15, which was slightly higher than the median value of calibrated X_5 in models MUOpt5H and MU6H (0.11 and 0.10, respectively). The similarity between calibrated X_5 and mean TIA was confirmed by the correlation coefficients, which reached 0.65–0.67 for models MUOpt5H and MU6H ($p < 0.001$). This suggests that calibrated X_5 acts as the catchment mean TIA, which can be viewed as a model-based assessment of extracted TIA from land cover. In addition, a greater improvement in calibration performances corresponded to a higher level of imperviousness. Unsurprisingly, the improvements in calibration performances were lower with very low X_5, i.e., as we converge to the setting of the original GR4H structure.

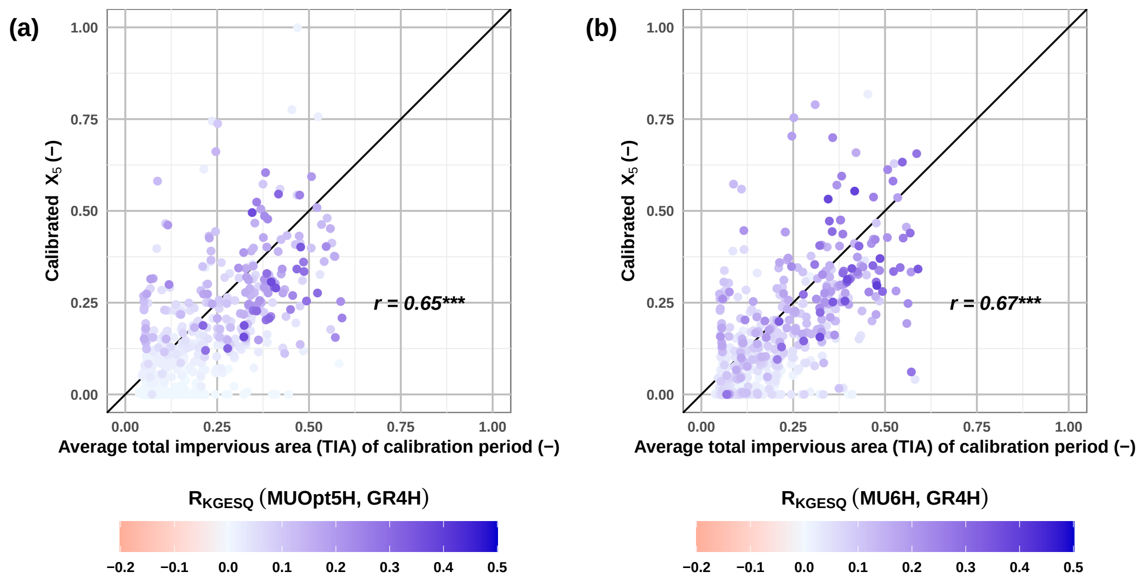

Figure 5. Calibrated proportion of impervious surfaces X_5 vs. estimated total impervious area from land cover databases. X_5 was calibrated for the models **(a)** MUOpt5H and **(b)** MU6H, as indicated in Table 2. Colors indicate the relative index R_{KGESQ}, i.e., the improvement in calibration performances with respect to the original GR4H structure. The black solid lines represent the identity lines. Asterisks *** indicate that the correlation coefficient r between calibrated X_5 and TIA was significant at a p-value threshold of 0.001.

The product of calibrated X_6 and X_5 (fixed at TIA or calibrated) had also high correlation values with TIA (0.72 to 0.81), as indicated in Figure 6. This high correlation was not surprising for MU5H as X_5 was constrained by TIA, while it suggests for MU6H that the product is highly related to urbanization. The fitted exponential functions were acceptable, and their corresponding coefficient of determination R^2 reached 0.42–0.58 when the scale factor was forced to 1 in order to obtain $X_5 \cdot X_6 = 1$ when $TIA = 1$ (dashed blue lines, Figures 6a and 6b). These curves had estimated exponents of $n = 1.86$ for MU5H (i.e., when $X_5 = TIA$) and $n = 2.62$ for MU6H (i.e., when X_5 was calibrated), and their comparison to the empirical, dashed grey curves of Sutherland (1995) suggests that the majority of catchments had behaviors similar to somewhat/extremely disconnected cases (last two bottom grey curves, Figures 6a and 6b). Calibrating both the scale factor and the exponent yielded equivalent R^2 values (red dashed lines, Figures 6a and 6b), with estimated parameters $m = 0.469$ and $n = 1.46$ for MU5H and $m = 0.687$ and $n = 2.45$ for MU6H. Still, the dispersion of $X_5 \cdot X_6$ may reflect the diversity of hydrological connectedness of impervious surfaces to the drainage network and/or the diversity of soil characteristics of the pervious part of the catchment.

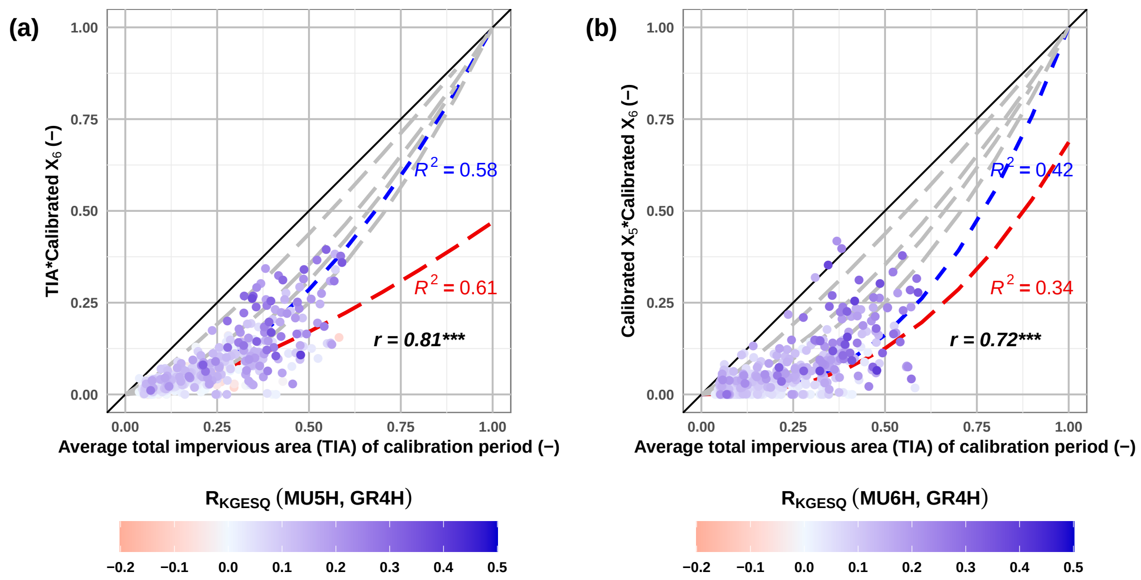

Figure 6. Product $X_5 \cdot X_6$ against mean total impervious area (TIA) per each calibration period for models **(a)** MU5H and **(b)** MU6H. X_5 represents the proportion of impervious surfaces in the catchment, which was fixed at TIA in MU5H and was calibrated in MU6H. X_6 is the quick-flow/slow-flow split parameter, which was calibrated for both models. Colors indicate the relative index R_{KGESQ}, i.e., the improvement in calibration performances relative to the original GR4H structure. The solid black lines represent the identity lines, and the grey dashed lines represent the Sutherland (1995) empirical relationships between TIA and effective impervious area (EIA), ranging from highly connected (i.e., close to the identity line) to extremely disconnected catchments (i.e., far from the identity line). Dashed blue lines represent the fitted relationships with scale factor forced to 1 (i.e., $X_5 \cdot X_6 = TIA^n$), and the dashed red lines represent the fitted relationships with calibrated scale factor and exponent (Equation 2.5). Asterisks *** indicate that the correlation coefficient r between $X_5 \cdot X_6$ and TIA was significant at a p-value threshold of 0.001.

3.2 Test performances

The tested modifications resulted in improved performances for the majority of catchments according to all continuous assessment criteria (Figure 7). The median relative index varied between 0.00 ($R_{KGESQ_{wet}}$ of MUOpt5H) and 0.08 (R_{NSE} of MU5H), corresponding to a median reduction in error ranging from 0% to 14.8%. The highest improvements were registered for NSE and $KGESQ_{dry}$, which suggests that the tested modifications were more beneficial for reproducing the high flows and the catchment response amid dry conditions. Exploiting the information of TIA to constrain the proportion of impervious surfaces X_5, with no additional calibrated parameters, gave better improvements than calibrating either X_5 or X_6 (MU4H vs. {MR5H, MUOpt5H}). The best overall improvements were obtained by MU5H and MU6H, whereas MUOpt5H showed fewer improvements, suggesting that a varying proportion of impervious surfaces X_5 should be accompanied by a varying quick-flow/slow-flow parameter X_6.

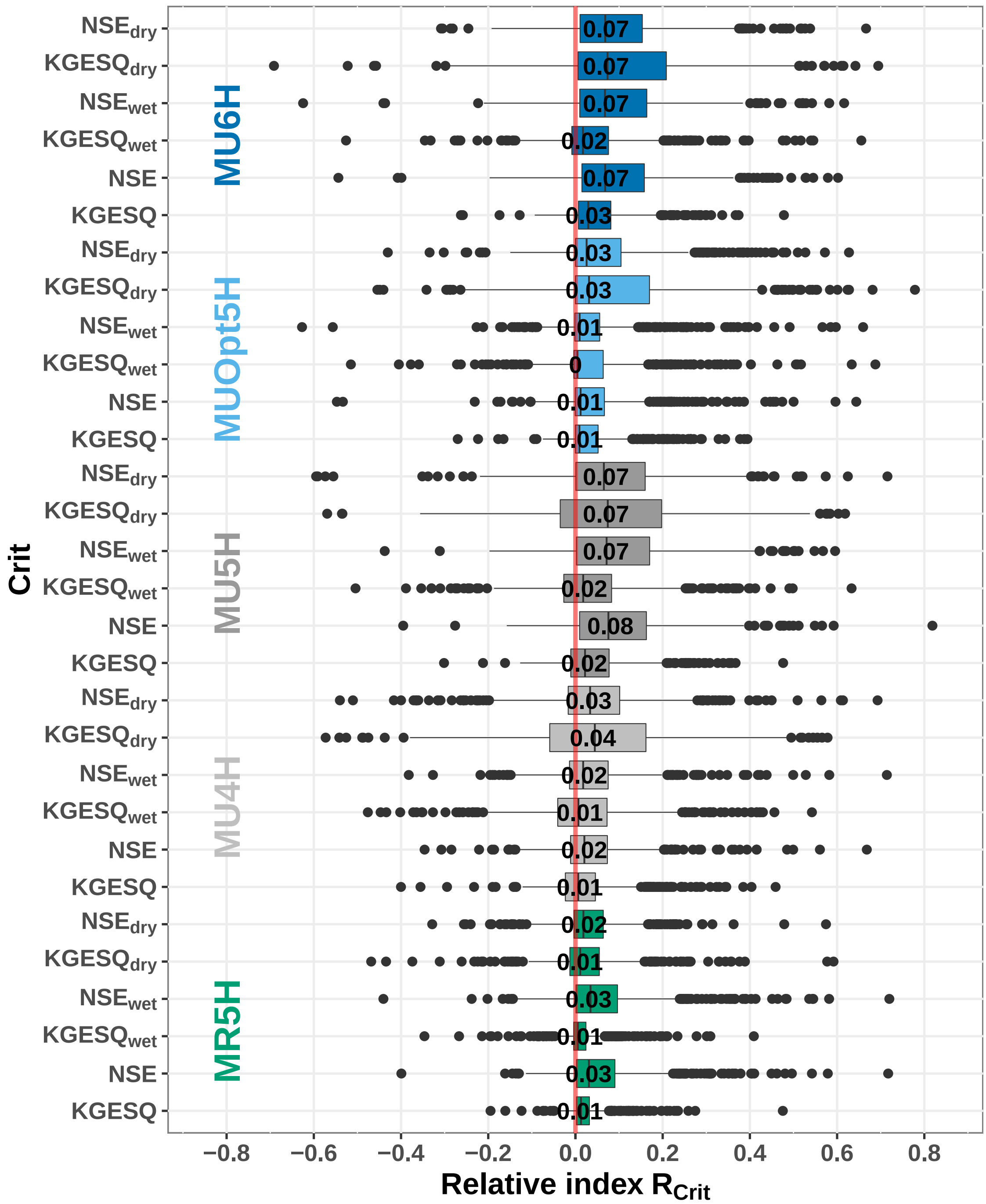

Figure 7. Distributions of improvements in test performances for the five tested modifications relative to the original model structure (GR4H), using six criteria for continuous assessment. Boxes are delimited by the first and third quartiles. Values indicate the median relative index R_{Crit} for each criterion $Crit$. The red line constitutes the border between improvement (on the right side) and deterioration (on the left side).

Concerning event-based assessment, the majority of modifications improved the performances over the 45,025 events, as suggested by the median values of event-based criteria shown in Table 3. In general, models tended to underestimate the event peak flow ($e_{Qp} < 0$). For GR4H, this underestimation reached on average 37.6% of the peak flow, but the error was reduced to roughly one fourth of the peak flow thanks to the

Table 3. Median values of event-based assessment criteria for the six tested models on 45,025 events. e_{Qp} indicates the error in estimating the event peak flow (ideal value: 0). e_{tp} indicates the error in estimating the timing of the event peak flow (ideal value: 0). VE is the volumetric efficiency, which is the error in estimating event runoff depths (ideal value: 1). The best performance for each criterion is highlighted in bold.

Criterion		Model					
		GR4H	**MR5H**	**MU4H**	**MU5H**	**MUOpt5H**	**MU6H**
e_{Qp}	Signed	−37.6%	−31.1%	−28.8%	**−23.6%**	−25.5%	−24.2%
	Absolute	51.3%	46.0%	45.1%	**39.7%**	44.2%	40.2%
e_{tp}	Signed	**0.0%**	−8.3%	**0.0%**	**0.0%**	**0.0%**	**0.0%**
	Absolute	17.9%	20.0%	**15.8%**	16.7%	16.7%	16.7%
	# of cases when $e_{tp} = 0$	11,925	10,721	**13,275**	12,007	12,580	11,808
VE		0.471	0.479	0.506	0.508	0.508	**0.509**

inclusion of an impervious surface proportion (fixed at TIA or calibrated, as in MU5H and MU6H). Absolute error in peak flow estimation using GR4H was approximately 51%, and was reduced to around 40% using MU5H or MU6H. The error in the timing of peak flow e_{tp} suggests that GR4H behaved as well as the modified versions, meaning that there was no significant improvement in predicting the timing of peak flow. The median absolute error was 17.9% of the delay between the beginning and the peak flow moments for GR4H, and these errors were slightly reduced to a median of 15.8% when the proportion of impervious surfaces was included. GR4H successfully predicted the timing of peak flow for 11,925 events (26.5% of events). This number was increased by MU4H (13,275), MU5H (12,007), and MUOpt5H (12,580), but they were all less than 30% of the total number of events. Finally, the median volumetric efficiency VE was improved by including the proportion of impervious surfaces, reaching a median value of 0.51 for all tested modifications, except for MR5H. This result confirms the usefulness of bypassing the soil moisture-accounting reservoir in order to better estimate event hydrographs.

The Wilcoxon signed-rank test suggested that all model modifications yielded statistically significant improvements in median values for each criterion ($p < 0.001$), except for MR5H, which had a significantly higher median of absolute errors in the estimation of the timing of peak flow relative to GR4H (not shown here). In addition, the binomial test indicated that all alternative versions were statistically better than the original model structure, except for the estimation of the timing of peak flow (Figure 8). As shown in Figure 7, the greatest improvements concerned $KGESQ$ in dry periods and NSE over the whole test period as well as in wet and dry conditions. Although the improvements in median $KGESQ$ were somewhat low, the number of cases at which this improvement was obtained was high, especially for MU6H and MR5H (85% and 83% of cases, respectively). Conversely, improvements in median NSE_{dry} were important, attaining an increase of +0.13 by MU5H, and NSE was improved for more than two thirds of the cases. A reduction in the error of peak flow estimation was obtained for nearly 59%–66% of events, whereas the volumetric efficiency was improved for 55%–60% of events. On the other hand, the tested modifications resulted in a similar estimation of the timing of the peak flow for the majority of events (range: 40%–59%), and improvements were registered only for approximately 26%–33% of events. The overall continuous and event-based performances are in favor of MU5H as an alternative model for urbanized catchments.

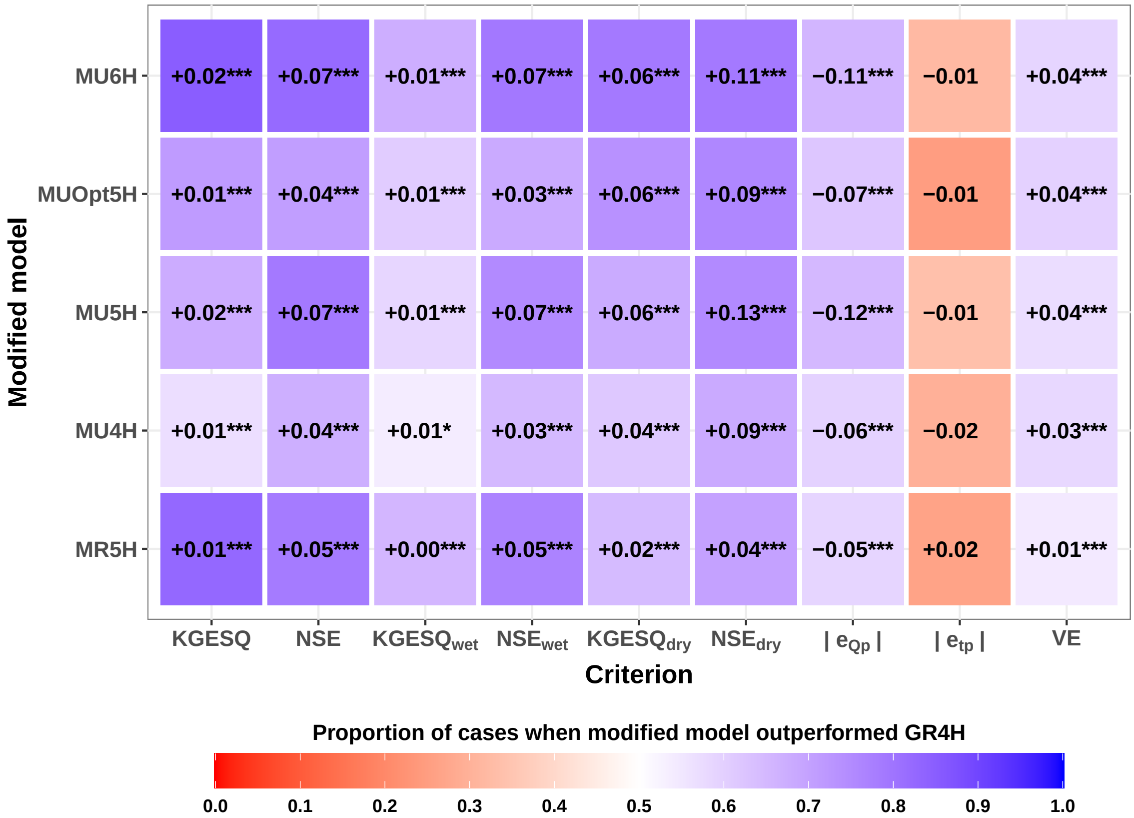

Figure 8. Summary of the improvements/deteriorations in median assessment criteria. Asterisks indicate the statistical significance of the resulting improvements using the binomial test, where *** indicates a statistical significance at a p-value threshold of 0.001 and * at a p-value threshold of 0.05. Colors indicate the proportion of cases when the modified model outperformed the original model GR4H. Note that improvements are obtained with positive differences in median values for all criteria except the absolute values of e_{Qp} and e_{tp}.

Qualitatively, Figure 9 suggests that the higher the urbanization level, the better the improvements by MU5H, as indicated by the distributions of the relative index R_{Crit} across TIA for the continuous assessment criteria. Improvements in $KGESQ$ were important only for TIA above 30%, whereas NSE was improved even for lower TIA levels and more importantly for higher TIA. Focusing on dry and wet conditions, improvements were less homogeneous, especially during dry conditions for all TIA levels. Overall, these results indicate that reproducing the catchment response was improved mostly for intensively urbanized catchments, and that during dry conditions, these improvements concerned also (but to a lesser extent) less urbanized cases.

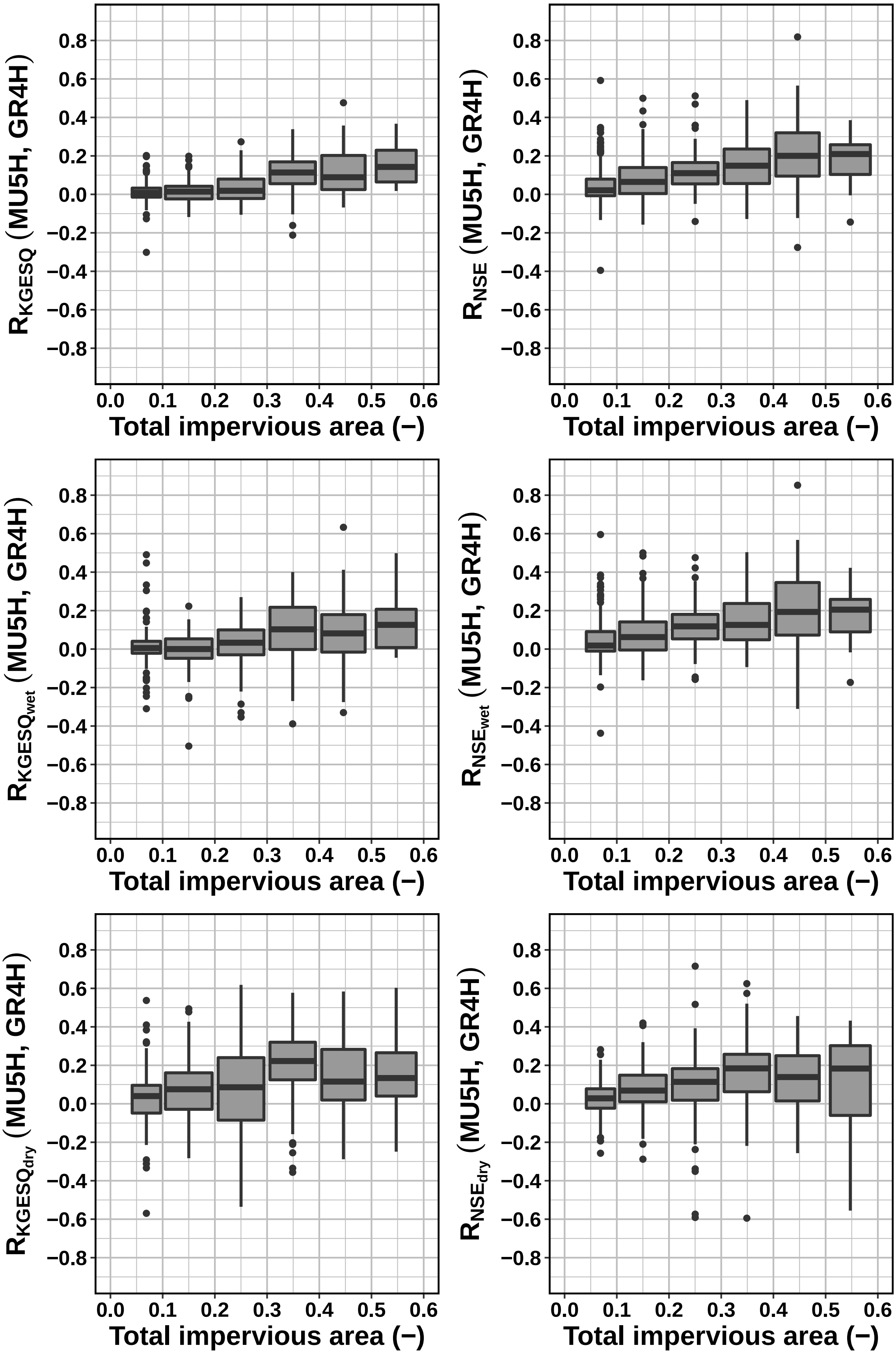

Figure 9. Improvements in reproducing the catchment response function of the total impervious area of the catchment during the test subperiod. Improvements are shown for the MU5H model, and are expressed using the relative index R_{Crit} (Equation 2.9).

4 Discussion and perspectives

4.1 Improvement of model performances

The attempted modifications aimed not only to enhance the link between model structure and urbanization features but also to improve the simulation of observed streamflow over the 273 urbanized catchments. Figures 7 and 8 show relatively limited but significant improvements in evaluation scores, which might not be convincing enough. First, a comparison with existing studies can guide the interpretation. Le Moine *et al.* (2007) compared a number of strategies to account for groundwater–surface water exchange at daily time steps within the GR4J model, and obtained an increase of +0.006 to +0.08 in NSE applied to square root values of simulated and observed streamflow depths, but for a larger number of 2080 cases of validation on 1040 catchments. Here, at the hourly time step, an improvement of +0.07 in NSE and +0.13 in NSE_{dry} was obtained for the 546 cases by MU5H. Similar improvements were obtained by Fidal and Kjeldsen (2020a), who accounted for infiltration in 28 urbanized sub-catchments of the Thames catchment (with proportion of urban surfaces between 1% and 55%) and obtained an increase in median NSE of +0.07. Another example was shown by Ficchì, Perrin, and Andréassian (2019), who obtained an improvement of +0.011 in median KGE on a sample of 240 catchments, which is lesser than the median improvements obtained here in terms of $KGESQ$. These comparisons show that the obtained improvements in the present study are relatively high. Second, these improvements were found statistically significant according to both the Wilcoxon test and the binomial test applied by Fidal and Kjeldsen (2020b), as shown in Figure 8. Interpreting these small improvements on large number of cases was difficult, it was thus necessary to apply statistical tests to make the claims robust enough.

Improvements were also obtained for event-based criteria but the errors in predicting the peak flow remained large even for the best alternative (MU5H, median absolute e_{Qp} = 39.7%). All models had a tendency to underestimate peak flows, which was also observed in previous model intercomparison studies (Boer-Euser *et al.*, 2017). We suspect that this is particularly attributed to the relatively small area of the urbanized catchments; 202 out of the 273 catchments had an area less than 150 km^2, and accounted for 81.6% of the total number of events. For similar drainage areas at the hourly time step, Lobligeois *et al.* (2014) had similar results and showed that better peak flow prediction was obtained on larger catchments. Another factor could be that the parameters were estimated without enough emphasis on high flows. We tested another set of parameters minimizing the KGE score on non-transformed streamflow values, but this did not yield substantially different results (not shown here).

Our evaluation procedure was statistical in the sense that we only focused on "average" performances or improvements, which overlooked some cases where model modifications degraded the performance relative to the original structure, as can be noticed in Figure 7. These cases suggest that adding more parameters in order to account for landscape specificities does not always lead to improved simulations of observed streamflows (Perrin, Michel, and Andréassian, 2001). Nonetheless, the improvements were obtained for a variety of climatic regions and levels of urbanization, implying the relevance of the modifications for a better conceptual understanding of the behavior of catchments with a mix of urban and non-urban surfaces. Thanks to these modifications, we obtained a robustly tested, continuum model between pervious and impervious areas, overcoming the dichotomous "urban–rural" conceptualization (Redfern *et al.*, 2016; McGrane, 2016; Brabec, Schulte, and Richards, 2002).

4.2 Improved linking of model structure to urbanization features

Although GR4H showed a high ability to reproduce the response of a large sample of urbanized catchments (Saadi, Oudin, and Ribstein, 2020b), its structure did not include any process specific to urban surfaces, such as the disconnection between soil moisture and runoff induced by high levels of imperviousness. This discouraged its use for predicting the impact of future urbanization on catchment behavior, obstructed by weakly sensitive parameters to urban measures (Saadi, Oudin, and Ribstein, 2019). The alternative models MU5H and MU6H preserve the simplicity and robustness of GR4H while introducing tightly related parameters to urban-specific characteristics, such as TIA and EIA.

We found that calibrated X_5 was highly correlated with mean TIA for the majority of cases (Figure 5). The coefficient of determination R^2 reached 0.45, which is high compared with reported results in regionalization studies (up to $R^2 = 0.27$ using 308 catchments by Merz and Blöschl, 2004; up to $R^2 = 0.61$ using 2225 catchments by Saadi, Oudin, and Ribstein, 2019). The product of X_5 (calibrated or fixed at TIA) and calibrated X_6 also showed significant correlations with TIA (Figure 6) and a pattern similar to classically adopted empirical equations describing the relationship between EIA and TIA (e.g., Sutherland, 1995). The scatter could be explained by the differences between catchments in terms of spatial arrangement of urban surfaces, their connectedness to the drainage network, or by the contrasts in soil and land-use characteristics of the pervious part of the catchment (Mejía and Moglen, 2010; Zhang and Shuster, 2014). Given the scale of the catchments we used in our study (1.1 km^2–2100 km^2) and their spatial heterogeneity, the obtained exponential relationships in Figure 6 may not be valid to estimate EIA from TIA at smaller scales (i.e., less than 1 km^2) or for cases with high values of TIA (i.e., higher than 0.6).

The use of MU5H to project the impact of urbanization on catchment behavior is possible by calibrating parameters X_1 to X_4 on a non-urbanized period of the catchment, constraining X_5 to TIA, and assigning X_6 depending on TIA and some knowledge concerning the connectedness of urban surfaces to the drainage system, aided by the relationship of Figure 6a. Conversely, calibrated X_5 and the product of calibrated X_5 with calibrated X_6 can be viewed as hydrologically estimated TIA and EIA using rainfall–runoff modeling and thus they could be used for a revision of estimated TIA and directly connected impervious area (DCIA) from land cover. This is a step forward towards conceiving hydrologically relevant measures of urbanization (Oudin *et al.*, 2018; Ebrahimian, Wilson, and Gulliver, 2016; Boyd, Bufill, and Knee, 1993; Alley and Veenhuis, 1983).

4.3 Limitations and perspectives

Our study was limited by the use of only one *a priori* estimated measure of urbanization (i.e., TIA), which has the merits of being simple to compute and relatively homogeneous over both France and the United States. This contributed to improve mainly the production component. Future model development should consider refining the routing component, where most of the controversies regarding urban impact lie (Oudin *et al.*, 2018), in relation to the effect of urbanization on low flows (Braud, Fletcher, and Andrieu, 2013). Potential modifications should enhance the representation of features related to subsurface flow and groundwater–surface water interactions in urbanized areas (as in Hamel and Fletcher, 2014). An attempt was illustrated by Furusho, Chancibault, and Andrieu (2013) by including the interaction between groundwater and sewerage systems in a distributed model, but a similar implementation would require a detailed description of the sewerage system of each catchment. Refining the routing component should also target on improving the event-based evaluation scores, particularly the error of peak flow estimation.

MU5H and MU6H have the merits of being robustly tested and relatively easy to implement, but their use is still less practical than other simple, conceptual models such as the CN method for the projection of the impact of urbanization. Newly added parameters X_5 and X_6 are prone to equifinality problems, their estimation from rainfall–runoff calibration could lead to their dependency on the climate characteristics of the calibration period (Brigode, Oudin, and Perrin, 2013; Merz, Parajka, and Blöschl, 2011) and on the chosen objective function. In addition, their calibrated values might reflect properties other than the ones they were a priori designed for (Andréassian *et al.*, 2012), such as runoff from soils of low permeability or catchment response to heavy rainfall. Hence, they should be better constrained using landscape measures from land cover databases (Gharari *et al.*, 2014). The empirical relationships of Figure 6 are a good starting point and they could be improved first by using information about DCIA or other measures of urban landscape fragmentation (Oudin *et al.*, 2018), and second by exploiting the characteristics of the pervious part of the catchment, namely its land-use and soil attributes. This will also help to link the model parameters to the spatial heterogeneities of the catchment, which in turn will help project the impact of different spatial arrangements of urban surfaces on catchment response (Zhang and Shuster, 2014; Mejía and Moglen, 2010). Future work should also address the impact of tested modifications on the parameters of the original structure GR4H (i.e., X_1 to X_4). Their interaction with X_5 and X_6 should be analyzed as well as their time instability and dependency on climate conditions. A strategy should be then proposed for a practical use of the alternative models MU5H and MU6H to project the impact of future urbanization and climate on catchment hydrology.

Acknowledgements

The authors thank the Handling Editor and the four anonymous reviewers for their constructive comments and suggestions which helped improve the manuscript.

References

Addor, N., *et al.* 2020. “Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges”. *Hydrological Sciences Journal* 65 (5): 712–725. ISSN: 0262-6667. doi:10.1080/02626667.2019.1683182. Cited on page 3.

Allen, R. G., *et al.* 1998. “Crop evapotranspiration - Guidelines for computing crop water requirements”. *FAO Irrigation and drainage paper* 56:15. Cited on page 5.

Alley, W. M., and J. E. Veenhuis. 1983. “Effective impervious area in urban runoff modeling”. *Journal of Hydraulic Engineering* 109 (2): 313–319. doi:10.1061/(ASCE)0733-9429(1983)109:2(313). Cited on pages 11, 24.

Andréassian, V., *et al.* 2009. “HESS Opinions "Crash tests for a standardized evaluation of hydrological models"”. *Hydrology and Earth System Sciences* 13 (10): 1757–1764. ISSN: 1027-5606. doi:10.5194/hess-13-1757-2009. Cited on pages 4, 13.

Andréassian, V., *et al.* 2012. “All that glitters is not gold: the case of calibrating hydrological models”. *Hydrological Processes* 26 (14): 2206–2210. ISSN: 1099-1085. doi:10.1002/hyp.9264. Cited on page 25.

Arnold, C. L. J., and C. J. Gibbons. 1996. “Impervious surface coverage: The emergence of a key environmental indicator”. *Journal of the American Planning Association* 62 (2): 243–258. ISSN: 0194-4363. doi:10.1080/01944369608975688. Cited on page 5.

Beck, H. E., *et al.* 2018. “Present and future Köppen-Geiger climate classification maps at 1-km resolution”. *Scientific Data* 5 (1): 1–12. ISSN: 2052-4463. doi:10.1038/sdata.2018.214. Cited on page 7.

Beven, K. J. 2002. “Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system”. *Hydrological Processes* 16 (2): 189–206. ISSN: 1099-1085. doi:10.1002/hyp.343. Cited on page 2.

Blume, T., E. Zehe, and A. Bronstert. 2007. “Rainfall—runoff response, event-based runoff coefficients and hydrograph separation”. *Hydrological Sciences Journal* 52 (5): 843–862. ISSN: 0262-6667. doi:10.1623/hysj.52.5.843. Cited on page 7.

Boer-Euser, T. de, *et al.* 2017. "Looking beyond general metrics for model comparison – lessons from an international model intercomparison study". *Hydrology and Earth System Sciences* 21 (1): 423–440. ISSN: 1027-5606. doi:10.5194/hess-21-423-2017. Cited on pages 7, 23.

Bonneau, J., *et al.* 2018. "The impact of urbanization on subsurface flow paths – A paired-catchment isotopic study". *Journal of Hydrology* 561:413–426. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2018.04.022. Cited on page 2.

Booth, D. B., and C. R. Jackson. 1997. "Urbanization of aquatic systems: Degradation thresholds, stormwater detection, and the limits of mitigation". *JAWRA Journal of the American Water Resources Association* 33 (5): 1077–1090. ISSN: 1752-1688. doi:10.1111/j.1752-1688.1997.tb04126.x. Cited on pages 5, 11.

Boyd, M. J., M. C. Bufill, and R. M. Knee. 1993. "Pervious and impervious runoff in urban catchments". *Hydrological Sciences Journal* 38 (6): 463–478. ISSN: 0262-6667. doi:10.1080/02626669309492699. Cited on page 24.

Brabec, E., S. Schulte, and P. L. Richards. 2002. "Impervious surfaces and water quality: A review of current literature and its implications for watershed planning". *Journal of Planning Literature* 16 (4): 499–514. ISSN: 0885-4122. doi:10.1177/088541202400903563. Cited on page 23.

Braud, I., T. D. Fletcher, and H. Andrieu. 2013. "Hydrology of peri-urban catchments: Processes and modelling". *Journal of Hydrology*, Hydrology of peri-urban catchments: processes and modelling, 485:1–4. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2013.02.045. Cited on page 24.

Braud, I., *et al.* 2013. "Evidence of the impact of urbanization on the hydrological regime of a medium-sized periurban catchment in France". *Journal of Hydrology*, Hydrology of peri-urban catchments: processes and modelling, 485:5–23. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2012.04.049. Cited on page 2.

Brigode, P., L. Oudin, and C. Perrin. 2013. "Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?" *Journal of Hydrology* 476:410–425. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2012.11.012. Cited on page 25.

Bronstert, A., D. Niehoff, and G. Bürger. 2002. "Effects of climate and land-use change on storm runoff generation: present knowledge and modelling capabilities". *Hydrological Processes* 16 (2): 509–529. ISSN: 1099-1085. doi:10.1002/hyp.326. Cited on page 2.

Cicco, L. A. D., *et al.* 2018. *dataRetrieval: R packages for discovering and retrieving water data available from U.S. federal hydrologic web services.* Reston, VA: U.S. Geological Survey. doi:10.5066/P9X4L3GE. Cited on page 5.

Clopper, C. J., and E. S. Pearson. 1934. "The use of confidence or fiducial limits illustrated in the case of the binomial". *Biometrika* 26 (4): 404–413. ISSN: 0006-3444. doi:10.1093/biomet/26.4.404. Cited on page 15.

Collischonn, W., and F. M. Fan. 2013. "Defining parameters for Eckhardt's digital baseflow filter". *Hydrological Processes* 27 (18): 2614–2622. ISSN: 1099-1085. doi:10.1002/hyp.9391. Cited on page 7.

Congedo, L., *et al.* 2016. "Copernicus high-resolution layers for land cover classification in Italy". *Journal of Maps* 12 (5): 1195–1205. doi:10.1080/17445647.2016.1145151. Cited on page 5.

Coron, L., *et al.* 2017. "The suite of lumped GR hydrological models in an R package". *Environmental Modelling and Software* 94:166–171. doi:10.1016/j.envsoft.2017.05.002. Cited on page 13.

Criss, R. E., and W. E. Winston. 2008. "Do Nash values have value? Discussion and alternate proposals". *Hydrological Processes* 22 (14): 2723–2725. ISSN: 1099-1085. doi:10.1002/hyp.7072. Cited on page 14.

Cristiano, E., M.-C. t. Veldhuis, and N. v. d. Giesen. 2017. "Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review". *Hydrology and Earth System Sciences* 21 (7): 3859–3878. ISSN: 1027-5606. doi:10.5194/hess-21-3859-2017. Cited on page 2.

Cuo, L., *et al.* 2008. "Hydrologic prediction for urban watersheds with the Distributed Hydrology–Soil–Vegetation Model". *Hydrological Processes* 22 (21): 4205–4213. ISSN: 1099-1085. doi:10.1002/hyp.7023. Cited on page 2.

De Niel, J., *et al.* 2020. "Efficient approach for impact analysis of land cover changes on hydrological extremes by means of a lumped conceptual model". *Journal of Hydrology: Regional Studies* 28:100666. ISSN: 2214-5818. doi:10.1016/j.ejrh.2020.100666. Cited on pages 2, 3.

Diem, J. E., T. C. Hill, and R. A. Milligan. 2018. "Diverse multi-decadal changes in streamflow within a rapidly urbanizing region". *Journal of Hydrology* 556:61–71. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2017.10.026. Cited on page 11.

Dotto, C. B. S., *et al.* 2011. "Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data". *Environmental Modelling & Software* 26 (10): 1225–1239. ISSN: 1364-8152. doi:10.1016/j.envsoft.2011.03.013. Cited on page 3.

Ebrahimian, A., B. N. Wilson, and J. S. Gulliver. 2016. "Improved methods to estimate the effective impervious area in urban catchments using rainfall-runoff data". *Journal of Hydrology* 536:109–118. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2016.02.023. Cited on pages 11, 24.

Eckhardt, K. 2005. "How to construct recursive digital filters for baseflow separation". *Hydrological Processes* 19 (2): 507–515. ISSN: 1099-1085. doi:10.1002/hyp.5675. Cited on page 7.

Edijatno *et al.* 1999. "GR3J: a daily watershed model with three free parameters". *Hydrological Sciences Journal* 44 (2): 263–277. ISSN: 0262-6667. doi:10.1080/02626669909492221. Cited on page 13.

Esse, W. R. van, *et al.* 2013. "The influence of conceptual model structure on model performance: a comparative study for 237 French catchments". *Hydrology and Earth System Sciences* 17 (10): 4227–4239. ISSN: 1027-5606. doi:10.5194/hess-17-4227-2013. Cited on pages 3, 7.

Euser, T., *et al.* 2015. "The effect of forcing and landscape distribution on performance and consistency of model structures". *Hydrological Processes* 29 (17): 3727–3743. ISSN: 1099-1085. doi:10.1002/hyp.10445. Cited on page 3.

Falcone, J. A. 2011. *GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow.* USGS Unnumbered Series. Reston, VA: U.S. Geological Survey. Visited on 07/23/2019. http://pubs.er.usgs.gov/publication/70046617. Cited on page 5.

Ficchì, A., C. Perrin, and V. Andréassian. 2019. "Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching". *Journal of Hydrology* 575:1308–1327. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2019.05.084. Cited on pages 3, 4, 7, 8, 9, 23.

– . 2016. "Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events". *Journal of Hydrology* 538:454–470. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2016.04.016. Cited on page 7.

Fidal, J., and T. R. Kjeldsen. 2020a. "Accounting for soil moisture in rainfall-runoff modelling of urban areas". *Journal of Hydrology* 589:125122. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2020.125122. Cited on pages 3, 23.

– . 2020b. "Operational comparison of rainfall-runoff models through hypothesis testing". *Journal of Hydrologic Engineering* 25 (4): 04020005. ISSN: 1943-5584. doi:10.1061/(ASCE)HE.1943-5584.0001892. Cited on pages 4, 15, 23.

Fletcher, T. D., H. Andrieu, and P. Hamel. 2013. "Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art". *Advances in Water Resources*, 35th Year Anniversary Issue, 51:261–279. ISSN: 0309-1708. doi:10.1016/j.advwatres.2012.09.001. Cited on pages 2, 4, 7.

Furusho, C., K. Chancibault, and H. Andrieu. 2013. "Adapting the coupled hydrological model ISBA-TOPMODEL to the long-term hydrological cycles of suburban rivers: Evaluation and sensitivity analysis". *Journal of Hydrology*, Hydrology of peri-urban catchments: processes and modelling, 485:139–147. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2012.06.059. Cited on page 24.

Gharari, S., *et al.* 2014. "Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration". *Hydrology and Earth System Sciences* 18 (12): 4839–4859. ISSN: 1027-5606. doi:10.5194/hess-18-4839-2014. Cited on pages 3, 4, 25.

Gupta, H. V., *et al.* 2014. "Large-sample hydrology: a need to balance depth with breadth". *Hydrology and Earth System Sciences* 18 (2): 463–477. ISSN: 1027-5606. doi:10.5194/hess-18-463-2014. Cited on pages 3, 4.

Gupta, H. V., T. Wagener, and Y. Liu. 2008. "Reconciling theory with observations: elements of a diagnostic approach to model evaluation". *Hydrological Processes* 22 (18): 3802–3813. ISSN: 1099-1085. doi:10.1002/hyp.6989. Cited on page 3.

Gupta, H. V., *et al.* 2009. "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling". *Journal of Hydrology* 377 (1): 80–91. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2009.08.003. Cited on page 13.

Haase, D. 2009. "Effects of urbanisation on the water balance – A long-term trajectory". *Environmental Impact Assessment Review* 29 (4): 211–219. ISSN: 0195-9255. doi:10.1016/j.eiar.2009.01.002. Cited on page 2.

Hamel, P., and T. D. Fletcher. 2014. "Modelling the impact of stormwater source control infiltration techniques on catchment baseflow". *Hydrological Processes* 28 (24): 5817–5831. ISSN: 1099-1085. doi:10.1002/hyp.10069. Cited on pages 3, 24.

Homer, C., *et al.* 2007. "Completion of the 2001 national land cover database for the counterminous United States". *Photogrammetric Engineering and Remote Sensing* 73 (4): 337–341. Cited on page 5.

Homer, C., *et al.* 2015. "Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a decade of land cover change information". *Photogrammetric Engineering & Remote Sensing* 81 (5): 345–354. ISSN: 0099-1112. doi:10.1016/S0099-1112(15)30100-2. Cited on page 5.

Homer, C., *et al.* 2004. "Development of a 2001 National Land-Cover Database for the United States". *Photogrammetric Engineering & Remote Sensing* 70 (7): 829–840. Cited on page 5.

Hrachowitz, M., *et al.* 2014. "Process consistency in models: The importance of system signatures, expert knowledge, and process complexity". *Water Resources Research* 50 (9): 7445–7469. ISSN: 1944-7973. doi:10.1002/2014WR015484. Cited on pages 3, 4.

Hrachowitz, M., and M. P. Clark. 2017. "HESS Opinions: The complementary merits of competing modelling philosophies in hydrology". *Hydrology and Earth System Sciences* 21 (8): 3953–3973. ISSN: 1027-5606. doi:10.5194/hess-21-3953-2017. Cited on page 2.

Huang, H.-j., *et al.* 2008. "Effect of growing watershed imperviousness on hydrograph parameters and peak discharge". *Hydrological Processes* 22 (13): 2075–2085. ISSN: 1099-1085. doi:10.1002/hyp.6807. Cited on page 3.

Jankowfsky, S., *et al.* 2014. "Assessing anthropogenic influence on the hydrology of small peri-urban catchments: Development of the object-oriented PUMMA model by integrating urban and rural hydrological models". *Journal of Hydrology* 517:1056–1071. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2014.06.034. Cited on page 2.

Jia, Y., *et al.* 2001. "Development of WEP model and its application to an urban watershed". *Hydrological Processes* 15 (11): 2175–2194. ISSN: 1099-1085. doi:10.1002/hyp.275. Cited on page 2.

Kirchner, J. W. 2006. "Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology". *Water Resources Research* 42 (3). ISSN: 1944-7973. doi:10.1029/2005WR004362. Cited on page 3.

Kjeldsen, T. R., J. D. Miller, and J. C. Packman. 2013. "Modelling design flood hydrographs in catchments with mixed urban and rural land cover". *Hydrology Research* 44 (6): 1040–1057. ISSN: 0029-1277. doi:10.2166/nh.2013.158. Cited on page 3.

Klemeš, V. 1986. "Operational testing of hydrological simulation models". *Hydrological Sciences Journal* 31 (1): 13–24. ISSN: 0262-6667. doi:10.1080/02626668609491024. Cited on pages 4, 13.

Krause, P., D. P. Boyle, and F. Bäse. 2005. "Comparison of different efficiency criteria for hydrological model assessment". In *Advances in Geosciences*, 5:89–97. Copernicus GmbH. doi:10.5194/adgeo-5-89-2005. Cited on page 13.

Langanke, T., *et al.* 2016. *Copernicus Land Monitoring Service – High Resolution Layer Imperviousness: Product Specifications Document*. Technical Report. European Environment Agency. https://land.copernicus.eu/user-corner/technical-library/hrl-imperviousness-technical-document-prod-2015. Cited on page 5.

Le Moine, N., *et al.* 2007. "How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments". *Water Resources Research* 43 (6). ISSN: 1944-7973. doi:10.1029/2006WR005608. Cited on pages 3, 7, 23.

Lee, J. G., and J. P. Heaney. 2003. "Estimation of urban imperviousness and its impacts on storm water systems". *Journal of Water Resources Planning and Management* 129 (5): 419–426. doi:10.1061/(ASCE)0733-9496(2003)129:5(419). Cited on page 11.

Legates, D. R., and G. J. McCabe. 1999. "Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation". *Water Resources Research* 35 (1): 233–241. ISSN: 1944-7973. doi:10.1029/1998WR900018. Cited on page 13.

Leleu, I., *et al.* 2014. "La refonte du système d'information national pour la gestion et la mise à disposition des données hydrométriques". *La Houille Blanche*, no. 1: 25–32. ISSN: 0018-6368, 1958-5551. doi:10.1051/lhb/2014004. Cited on page 5.

Leopold, L. B. 1968. "Hydrology for urban land planning: A guidebook on the hydrologic effects of urban land use". *Geological Survey Circular* 554. Cited on pages 2, 9.

Lerat, J., *et al.* 2012. "Do internal flow measurements improve the calibration of rainfall-runoff models?" *Water Resources Research* 48 (2). ISSN: 1944-7973. doi:10.1029/2010WR010179. Cited on page 14.

Lin, Y., and K. E. Mitchell. 2005. "The NCEP stage II/IV hourly precipitation analyses: Development and applications". In *Proceedings of the 19th conference on Hydrology*, 9–13. Citeseer. Cited on page 5.

Lobligeois, F., *et al.* 2014. “When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events”. *Hydrology and Earth System Sciences* 18 (2): 575–594. ISSN: 1607-7938. doi:10.5194/hess-18-575-2014. Cited on pages 7, 23.

McGrane, S. J. 2016. “Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review”. *Hydrological Sciences Journal* 61 (13): 2295–2311. ISSN: 0262-6667. doi:10.1080/02626667.2015.1128084. Cited on pages 2, 23.

McIntyre, N., *et al.* 2014. “Modelling the hydrological impacts of rural land use change”. *Hydrology Research* 45 (6): 737–754. ISSN: 0029-1277. doi:10.2166/nh.2013.145. Cited on page 2.

McMillan, H. K., *et al.* 2011. “Hydrological field data from a modeller’s perspective: Part 1. Diagnostic tests for model structure”. *Hydrological Processes* 25 (4): 511–522. ISSN: 1099-1085. doi:10.1002/hyp.7841. Cited on page 3.

Mei, Y., and E. N. Anagnostou. 2015. “A hydrograph separation method based on information from rainfall and runoff records”. *Journal of Hydrology* 523:636–649. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2015.01.083. Cited on page 7.

Mejía, A. I., and G. E. Moglen. 2010. “Impact of the spatial distribution of imperviousness on the hydrologic response of an urbanizing basin”. *Hydrological Processes* 24 (23): 3359–3373. ISSN: 1099-1085. doi:10.1002/hyp.7755. Cited on pages 24, 25.

Merz, R., and G. Blöschl. 2004. “Regionalisation of catchment model parameters”. *Journal of Hydrology* 287 (1): 95–123. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2003.09.028. Cited on page 24.

Merz, R., J. Parajka, and G. Blöschl. 2009. “Scale effects in conceptual hydrological modeling”. *Water Resources Research* 45 (9). ISSN: 1944-7973. doi:10.1029/2009WR007872. Cited on page 5.

Merz, R., J. Parajka, and G. Blöschl. 2011. “Time stability of catchment model parameters: Implications for climate impact analyses”. *Water Resources Research* 47 (2). ISSN: 1944-7973. doi:10.1029/2010WR009505. Cited on page 25.

Michel, C., V. Andréassian, and C. Perrin. 2005. “Soil Conservation Service Curve Number method: How to mend a wrong soil moisture accounting procedure?” *Water Resources Research* 41 (2). ISSN: 1944-7973. doi:10.1029/2004WR003191. Cited on page 3.

Miller, J. D., and T. Hess. 2017. “Urbanisation impacts on storm runoff along a rural-urban gradient”. *Journal of Hydrology* 552:474–489. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2017.06.025. Cited on pages 2, 9, 11.

Nash, J. E., and J. V. Sutcliffe. 1970. “River flow forecasting through conceptual models part I — A discussion of principles”. *Journal of Hydrology* 10 (3): 282–290. ISSN: 0022-1694. doi:10.1016/0022-1694(70)90255-6. Cited on page 13.

Niehoff, D., U. Fritsch, and A. Bronstert. 2002. “Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany”. *Journal of Hydrology*, Advances in Flood Research, 267 (1): 80–93. ISSN: 0022-1694. doi:10.1016/S0022-1694(02)00142-7. Cited on page 2.

Ogden, F. L., *et al.* 2017. “Comment on “Beyond the SCS-CN method: A theoretical framework for spatially lumped rainfall-runoff response” by M. S. Bartlett et al.” *Water Resources Research* 53 (7): 6345–6350. ISSN: 1944-7973. doi:10.1002/2016WR020176. Cited on page 3.

Ogden, F. L., *et al.* 2011. “Relative importance of impervious area, drainage density, width function, and subsurface storm drainage on flood runoff from an urbanized catchment”. *Water Resources Research* 47 (12). ISSN: 1944-7973. doi:10.1029/2011WR010550. Cited on page 2.

Oudin, L., *et al.* 2006. “Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations”. *Water Resources Research* 42 (7). ISSN: 1944-7973. doi:10.1029/2005WR004636. Cited on pages 13, 14.

Oudin, L., *et al.* 2018. “Hydrological impacts of urbanization at the catchment scale”. *Journal of Hydrology* 559:774–786. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2018.02.064. Cited on pages 2, 24, 25.

Oudin, L., *et al.* 2005. “Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling”. *Journal of Hydrology* 303 (1): 290–306. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2004.08.026. Cited on page 5.

Pathiraja, S., *et al.* 2018. “Time-varying parameter models for catchments with land use change: the importance of model structure”. *Hydrology and Earth System Sciences* 22 (5): 2903–2919. ISSN: 1027-5606. doi:10.5194/hess-22-2903-2018. Cited on page 2.

Perrin, C., C. Michel, and V. Andréassian. 2001. "Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments". *Journal of Hydrology* 242 (3): 275–301. ISSN: 0022-1694. doi:10.1016/S0022-1694(00)00393-0. Cited on pages 3, 23.

Perrin, C., C. Michel, and V. Andréassian. 2003. "Improvement of a parsimonious model for streamflow simulation". *Journal of Hydrology* 279 (1): 275–289. ISSN: 0022-1694. doi:10.1016/S0022-1694(03)00225-7. Cited on pages 8, 9, 11.

Perrin, C., *et al.* 2007. "Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models". *Hydrological Sciences Journal* 52 (1): 131–151. ISSN: 0262-6667. doi:10.1623/hysj.52.1.131. Cited on page 5.

Petrucci, G., and C. Bonhomme. 2014. "The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data". *Journal of Hydrology* 517:997–1007. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2014.06.019. Cited on page 2.

Prosdocimi, I., T. R. Kjeldsen, and J. D. Miller. 2015. "Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models". *Water Resources Research* 51 (6): 4244–4262. ISSN: 1944-7973. doi:10.1002/2015WR017065. Cited on page 2.

Read, J. S., *et al.* 2015. "geoknife: Reproducible web-processing of large gridded datasets". *Ecography.* doi:10.1111/ecog.01880. Cited on page 5.

Redfern, T. W., *et al.* 2016. "Current understanding of hydrological processes on common urban surfaces". *Progress in Physical Geography: Earth and Environment* 40 (5): 699–713. ISSN: 0309-1333. doi:10.1177/0309133316652819. Cited on pages 2, 4, 23.

Rodriguez, F., H. Andrieu, and J.-D. Creutin. 2003. "Surface runoff in urban catchments: morphological identification of unit hydrographs from urban databanks". *Journal of Hydrology* 283 (1): 146–168. ISSN: 0022-1694. doi:10.1016/S0022-1694(03)00246-4. Cited on pages 2, 7.

Saadi, M., L. Oudin, and P. Ribstein. 2020a. "Beyond imperviousness: The role of antecedent wetness in runoff generation in urbanized catchments". *Water Resources Research* 56 (11): e2020WR028060. ISSN: 1944-7973. doi:https://doi.org/10.1029/2020WR028060. Cited on pages 5, 7, 9.

– . 2020b. "Crossing the rural–urban boundary in hydrological modelling: How do conceptual rainfall–runoff models handle the specificities of urbanized catchments?" *Hydrological Processes* 34 (15): 3331–3346. ISSN: 1099-1085. doi:10.1002/hyp.13808. Cited on pages 2, 3, 4, 8, 11, 24.

– . 2019. "Random forest ability in regionalizing hourly hydrological model parameters". *Water* 11 (8): 1540. doi:10.3390/w11081540. Cited on pages 4, 24.

Salavati, B., *et al.* 2016. "Modeling approaches to detect land-use changes: Urbanization analyzed on a set of 43 US catchments". *Journal of Hydrology* 538:138–151. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2016.04.010. Cited on page 2.

Salvadore, E., J. Bronders, and O. Batelaan. 2015. "Hydrological modelling of urbanized catchments: A review and future directions". *Journal of Hydrology* 529:62–81. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2015.06.028. Cited on pages 2, 3, 4, 5, 7.

Santos, L., G. Thirel, and C. Perrin. 2018. "Technical note: Pitfalls in using log-transformed flows within the KGE criterion". *Hydrology and Earth System Sciences* 22 (8): 4583–4591. ISSN: 1027-5606. doi:10.5194/hess-22-4583-2018. Cited on page 13.

Sanzana, P., *et al.* 2019. "Impact of urban growth and high residential irrigation on streamflow and groundwater levels in a peri-urban semiarid catchment". *JAWRA Journal of the American Water Resources Association* 55 (3): 720–739. ISSN: 1752-1688. doi:10.1111/1752-1688.12743. Cited on page 2.

Schaefli, B., and H. V. Gupta. 2007. "Do Nash values have value?" *Hydrological Processes* 21 (15): 2075–2080. ISSN: 1099-1085. doi:10.1002/hyp.6825. Cited on page 13.

Sillanpää, N., and H. Koivusalo. 2015. "Impacts of urban development on runoff event characteristics and unit hydrographs across warm and cold seasons in high latitudes". *Journal of Hydrology* 521:328–340. ISSN: 0022-1694. doi:10.1016/j.jhydrol.2014.12.008. Cited on page 14.

Stavropulos-Laffaille, X., *et al.* 2018. "Improvements to the hydrological processes of the Town Energy Balance model (TEB-Veg, SURFEX v7.3) for urban modelling and impact assessment". *Geoscientific Model Development* 11 (10): 4175–4194. ISSN: 1991-959X. doi:10.5194/gmd-11-4175-2018. Cited on page 2.

Sutherland, R. C. 1995. "Methodology for estimating the effective impervious area of urban watersheds". *Watershed Protection Techniques* 2 (1): 282–284. Cited on pages 11, 17, 18, 24.

Tabary, P., *et al.* 2012. "A 10-year (1997–2006) reanalysis of Quantitative Precipitation Estimation over France: methodology and first results". *IAHS Publication* 351:255–260. Cited on page 5.

Terrier, M., *et al.* 2021. "Streamflow naturalization methods: a review". *Hydrological Sciences Journal* 66 (1): 12–36. ISSN: 0262-6667. doi:10.1080/02626667.2020.1839080. Cited on page 5.

Thornton, P. E., *et al.* 2016. "Daymet: Daily surface weather data on a 1-km grid for North America, version 3". *ORNL DAAC.* doi:10.3334/ORNLDAAC/1328. Cited on page 5.

Vidal, J.-P., *et al.* 2010. "A 50-year high-resolution atmospheric reanalysis over France with the Safran system". *International Journal of Climatology* 30 (11): 1627–1644. ISSN: 1097-0088. doi:10.1002/joc.2003. Cited on page 5.

Walsh, C. J., *et al.* 2005. "The urban stream syndrome: current knowledge and the search for a cure". *Journal of the North American Benthological Society* 24 (3): 706–723. ISSN: 0887-3593. doi:10.1899/04-028.1. Cited on page 9.

Wickham, J., *et al.* 2014. "The Multi-Resolution Land Characteristics (MRLC) Consortium — 20 years of development and integration of USA National Land Cover Data". *Remote Sensing* 6 (8): 7424–7441. doi:10.3390/rs6087424. Cited on page 5.

Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods". *Biometrics Bulletin* 1 (6): 80–83. ISSN: 0099-4987. doi:10.2307/3001968. Cited on page 15.

Zhang, Y., and W. Shuster. 2014. "Impacts of spatial distribution of impervious areas on runoff response of hillslope catchments: Simulation study". *Journal of Hydrologic Engineering* 19 (6): 1089–1100. ISSN: 1943-5584. doi:10.1061/(ASCE)HE.1943-5584.0000905. Cited on pages 24, 25.

Zhou, Z., *et al.* 2017. "The complexities of urban flood response: Flood frequency analyses for the Charlotte metropolitan region". *Water Resources Research* 53 (8): 7401–7425. ISSN: 1944-7973. doi:10.1002/2016WR019997. Cited on pages 9, 14.