Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR024143

The Influence of Rainfall and Catchment Critical Scales on Urban Hydrological Response Sensitivity

Elena Cristiano[1], Marie-claire ten Veldhuis[1], Daniel B. Wright[2], James A. Smith[3], and Nick van de Giesen[1]

[1]Department of Water Management, Delft University of Technology, Delft, The Netherlands, [2]Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA, [3]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

Key Points:
- Rainfall aggregation has high impact on the hydrological response in urban areas
- Characterization of rainfall spatial and temporal scales plays a significant role in the estimation of hydrological response sensitivity
- Combination of rainfall and catchment critical scales strongly affects the sensitivity of the hydrological response

Supporting Information:
- Supporting Information S1

Correspondence to:
E. Cristiano,
e.cristiano@tudelft.nl

Citation:
Cristiano, E., ten Veldhuis, M.-C., Wright, D. B., Smith, J. A., & van de Giesen, N. (2019). The influence of rainfall and catchment critical scales on urban hydrological response sensitivity. *Water Resources Research*, *55*, 3375–3390. https://doi.org/10.1029/2018WR024143

Received 23 SEP 2018
Accepted 20 MAR 2019
Accepted article online 11 APR 2019
Published online 29 APR 2019

Abstract Interactions between spatial and temporal variability of rainfall and catchment characteristics strongly influence hydrological response. In urban areas, where runoff generation is fast due to high imperviousness degree, it is especially relevant to capture the high spatiotemporal rainfall variability. Significant progress has been made in the development of spatially distributed rainfall measurements and of distributed hydrological models, to represent the variability of catchment's characteristics. Interactions between rainfall and basin scales on hydrological response sensitivity, however, needs deeper investigation. A previous study investigated the hydrological response in the small urbanized catchment of Cranbrook (8 km^2, London, UK) and proposed three dimensionless "scale factors" to identify if the available rainfall resolution is sufficient to properly predict hydrological response. We aim to verify the applicability of these scale factors to larger scales, with a distinct physiographic setting, in Little Sugar Creek (111 km^2, Charlotte, USA), to identify the required rainfall resolution and to predict model performance. Twenty-eight events were selected from a weather radar data set from the National Weather Radar Network, with a resolution of 1 km^2 and 15 min. Rainfall data were aggregated to coarser resolutions and used as input for a distributed hydrological model. Results show that scale factors and associated thresholds are generally applicable for characterization of urban flood response to rainfall across spatiotemporal scales. Additionally, application of scale factors in observation-based analysis supports identification of event characteristics that are poorly captured and critical improvements that need to be made before the model can benefit from high-resolution rainfall.

1. Introduction

Hydrological response is strongly influenced by interactions between rainfall variability in space and time and catchment characteristics. These interactions are particularly pronounced in urban areas, where the runoff generation is fast, due to the high degree of imperviousness and to the large heterogeneity of catchment characteristics. For these reasons, the use of high-resolution rainfall data is necessary to investigate and predict hydrological response in urban systems (Faures et al., 1995; Ochoa-Rodriguez et al., 2015; Sempere-Torres et al., 1999; Smith et al., 2012; Zhou et al., 2017). In recent decades, new technologies and instruments have been developed to measure rainfall variability in space and time at high resolution (Leijnse et al., 2007, 2010; van de Beek et al., 2010). Rainfall data derived from weather radar in particular are increasingly used for urban hydrological applications (Einfalt et al., 2004; Thorndahl et al., 2017). Although weather radars provide indirect rainfall measurements and require calibration with ground observations, they provide spatially distributed rainfall estimates, which are essential for understanding the effects of rainfall on hydrological response (Berne et al., 2004; Niemczynowicz, 1988; Ochoa-Rodriguez et al., 2015; Rafieeinasab et al., 2015; Schilling, 1991; Yang et al., 2016).

Thanks to increasing computational power and the availability of high-resolution geomorphological data (Tokarczyk et al., 2015), high-resolution distributed models have been developed and implemented for urban areas, where detailed representation of surface and drainage network is especially important (Gironás et al., 2010; Pina et al., 2016; Smith et al., 2013). Poor model performance can arise from inadequate model structure and representation of catchment and network characteristics, in addition to rainfall errors and low resolution (Pina et al., 2016; Wright et al., 2014a). Models can represent the surface in different ways, including lumped, semidistributed using lumped subcatchments and fully distributed schemes, in which the

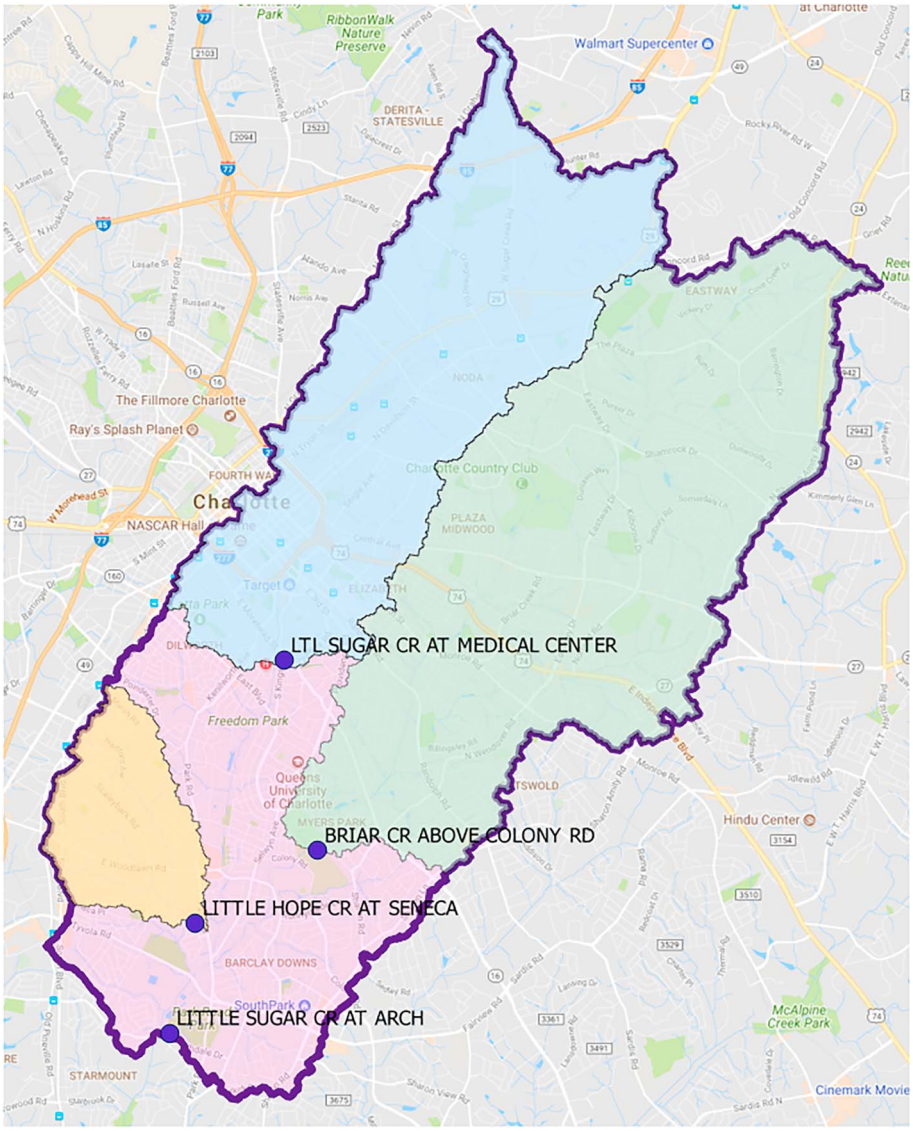

Figure 1. Map of the selected catchments and locations.

surface is represented with a regular or irregular mesh (Salvadore et al., 2015; Zoppou, 2000). The choice of model type that is best suited for a given application can be strongly influenced by the resolution of available rainfall data (Pina et al., 2014, 2016).

With the emergence of high-resolution models and rainfall observations, relationships between rainfall and catchment scales and their impact on hydrological response prediction can now be investigated in detail (Berne et al., 2004; Bruni et al., 2015; Cristiano et al., 2017; Ochoa-Rodriguez et al., 2015; ten Veldhuis et al., 2018).

Ochoa-Rodriguez et al. (2015), following the study presented by Berne et al. (2004), showed that there is a relation between increasing drainage area and decreasing runoff sensitivity to coarser rainfall resolutions. Their results, based on hydrological simulations of seven different catchments, allowed to derive several scale factors, that combine rainfall scale with the minimum required rainfall resolution to investigate the influence of rainfall characteristics on the hydrological response, and they highlighted the importance of combining both spatial and temporal rainfall characteristics.

In a follow-up study by Cristiano et al. (2018), three dimensionless parameters were introduced to represent the interactions between rainfall and basin scale and their effects on hydrological response. These

Table 1
Characteristics of the Selected Subcatchments

ID	Name	Connected area (km^2)	Imperviousness (%)	Slope (%)
02146507	Little Sugar Creek at Archdale	111	32	2.4
02146470	Little Hope Creek	7	32.2	2.2
(02146)45022	Briar Creek	48.5	24.7	2.4
(02146)409	Little Sugar Creek at Med. Centre	31	48.2	2.2

Note. Catchment IDs follow the U.S. Geological Survey classification.

dimensionless factors combine spatial and temporal rainfall scales and spatial and temporal catchment characteristic scales in relation with the rainfall spatial resolution Δs and temporal resolution Δt. The performance of the hydrological model was evaluated in relation to a range of scale factor values, from which thresholds associated with certain minimum levels of model performance were derived. These scale factors and related thresholds were developed for the dense-urbanized small basin (9 km^2) of Cranbrook (London, UK).These factors and metrics are tested here for very different climatological and geomorphological catchment conditions and at a larger catchment scale. The aim of this study is to investigate the applicability of the scale factors and performance thresholds for catchments in the Charlotte metropolitan region (NC, USA).

The paper is structured as follows. Rainfall data, catchment, and model description are presented in section 2. Section 3 describes the methodology used in the current study to investigate the effects of rainfall aggregation, scale factors, and their applicability. Section 4 shows the results obtained in the study. Section 5 summarizes the main findings of the research, focusing on the analysis of the rainfall event characteristics that affect hydrologic response sensitivity.

2. Data

2.1. Catchment and Model Description

The Charlotte metropolitan region (North Carolina, USA) is largely located within the watershed of Little Sugar Creek (111 km^2). As shown in Figure 1, four subcatchments of Little Sugar Creek, spanning a range of drainage areas and degrees of imperviousness (Table 1) were selected for this study to investigate relationships between rainfall resolutions and hydrological response across a range of catchment scales. Catchment dimensions vary form 7 to 100 km^2. Flow measurements at the outlet of each subcatchment at 1-min resolution are used as reference when investigating hydrological response sensitivity, in the second part of the this study (section 4.3.2). Flood response in Little Sugar Creek were at first investigated by Smith et al. (2002). In this work, additional information about catchment characteristics and rainfall variability in this area can be found.

The physically based distributed Gridded Surface Subsurface Hydrologic Analysis (GSSHA, for a complete description see Downer & Ogden, 2003, 2004) model was used to estimate the hydrological response of the system. A GSSHA model, including river network and major subsurface storm drain in the Little Sugar Creek catchment, was developed and used in previous studies (Wright et al., 2014a). The GSSHA model uses explicit finite-volume solutions of the diffusive wave formulation of the St. Venant 2-D equations to simulate overland flow in the grid and 1-D equations for channel flow. High-resolution geographic information system-based data sets were used as input for the model. Groundwater and subsurface runoff production were neglected, assuming that the Hortonian runoff is predominant in urban areas, due to the high imperviousness degree. Among the different infiltration models available with GSSHA, a three-layer Green-Ampt infiltration scheme was chosen to represent the infiltration processes and the percentage of permeable grid cells was defined according to imperviousness degree derived from the cover map from the 2006 National Land Cover Data set (NLCD) data set. The original GSSHA source code was modified in order to enable the Manning's roughness coefficient for 1-D channel flow to vary across the width of a channel cross section. Storm drains with diameter smaller than 0.91 were not included. The model represents the basin using a 90 $\times$ 90 m^2 grid and stream flow was simulated with a temporal resolution of 1 min.

The GSSHA model for Little Sugar Creek was validated using U.S. Geological Survey streamflow observations at the outlets of the four investigated subcatchments for 10 rainfall events. Results showed that the

Table 2
Characteristics of the Selected Rainfall Events

ID	Starting day UTC	Starting time UTC	Duration (hr)	Maximum intensity (mm/15 min)	Total depth (mm)
Ev1	2001-07-18	00:15	4.75	14.04	35.01
Ev2	2002-07-14	00:00	15.75	11.97	34.36
Ev3	2002-08-16	19:00	12.75	19.06	54.53
Ev4	2003-05-21	15:15	38.25	11.19	65.90
Ev5	2003-06-16	21:15	6	17.46	33.80
Ev6	2003-06-18	01:30	2.75	21.31	12.65
Ev7	2003-07-19	22:00	3.5	25.95	28.89
Ev8	2003-08-04	21:30	8.25	17.27	29.44
Ev9	2004-08-12	11:15	18.5	10.37	30.53
Ev10	2004-09-07	09:00	26.75	9.76	59.47
Ev11	2005-05-10	20:00	10.75	25.95	47.73
Ev12	2005-06-07	20:00	4.75	25.95	28.00
Ev13	2005-07-07	17:30	4.5	12.23	14.48
Ev14	2005-08-23	17:15	3	18.89	18.57
Ev15	2006-08-15	23:30	16.5	25.93	93.85
Ev16	2007-09-13	02:00	27.75	2.98	37.93
Ev17	2008-09-10	14:00	14.75	24.50	34.82
Ev18	2009-05-05	19:30	9.5	5.69	8.31
Ev19	2009-05-25	13:45	14.75	18.01	19.01
Ev20	2009-06-04	13:30	26.75	6.11	38.61
Ev21	2009-08-16	16:30	5.5	24.01	44.79
Ev22	2010-08-19	05:00	21.75	18.69	55.44
Ev23	2011-08-05	13:30	12.75	22.57	77.52
Ev24	2011-09-22	23:45	16.25	1.76	7.86
Ev25	2012-07-12	17:15	4.25	22.07	47.76
Ev26	2013-06-07	23:00	10.25	20.86	36.10
Ev27	2013-09-21	15:00	14.75	3.17	17.85
Ev28	2015-04-19	08:30	16.5	11.69	26.66

Note. Total rainfall depth is accumulated over the whole drainage area of Little Sugar Creek. Dates are formatted as YYYY-MM-DD.

central tendency of flood response is captured by the model, although some systematic bias in streamflow exists in modeled peak discharge. Bias may be attributable to errors in radar rainfall estimates or to features that were omitted in the model (Wright et al., 2014b). A complete description of model characteristics and validation can be found in Wright et al. (2014a).

2.2. Rainfall Events

Twenty-eight storm events were selected from a 15-year (2001–2015) high-resolution (15 min, 1 km^2) radar data set (Wright et al., 2014b), developed from bias-corrected radar reflectivity observations from the National Weather Service (NWS) Next Generation Radar network (NEXRAD) processed using the Hydro-NEXRAD algorithm (Krajewski et al., 2007; Seo et al., 2010). Quality control algorithms and Z-R conversion of reflectivity to rainfall rate are included in the Hydro-NEXRAD system. Rainfall fields derived with this algorithm have been used as input for hydrological analysis in several studies (Seo et al., 2010; ten Veldhuis & Schleiss, 2017; ten Veldhuis et al., 2018; Wright et al., 2013, 2014b; Zhou et al., 2017)

Table 2 summarizes rainfall event characteristics, presenting starting time, event duration, maximum intensity in a single pixel, and total depth. Two events are assumed to be independent when there is at

least a 6-hr dry period between them, following the most commonly approach found in the literature (Driscoll et al., 1989; Restrepo-Posada & Eagleson, 1982; Xuereb & Green, 2012). The approach proposed by Restrepo-Posada and Eagleson (1982) was evaluated for the United Stated in Driscoll et al. (1989), showing how two events can be considered independent if there is a minimum dry period of 6 hr. Events present different characteristics, with durations varying between 3 and 38 hr and rainfall peak intensities between 8 and 104 mm/hr. Hietographs and hydrographs of the selected rainfall events are presented in the supporting information.

3. Method

3.1. Rainfall Aggregation

To investigate hydrological response sensitivity to different spatial and temporal rainfall resolutions, radar rainfall fields were aggregated in space (to 3×3 and 6×6 km^2) and time (to 30 and 60 min). Interpolation techniques to obtain a finest rainfall resolutions (Ferraris et al., 2003; Gires et al., 2012; Muthusamy et al., 2017) were not applied in this study, since they would inevitably introduce uncertainties in the rainfall estimates that could mask the effects we aimed to study.

Rainfall events at these different resolutions were used as input for the hydrological model to investigate the effects of aggregation on hydrological response. Simulated stream flows at subcatchment outlets, generated using coarser rainfall resolutions, were first compared with the simulated streamflow obtained using the original 15 min-1 km^2 rainfall resolution as input of the model, and later with observed flows.

Effects of rainfall aggregation are investigated using the peak attenuation ratio P proposed by Cristiano et al. (2018) and defined as follows:

$$P = -\frac{R_{\max,\mathrm{ref}} - R_{\max,\Delta s \Delta t}}{R_{\max,\mathrm{ref}}} \tag{1}$$

where $R_{\max,\mathrm{ref}}$ is the rainfall peak rate of single pixel at 1 km^2-15 min and $R_{\max,\mathrm{st}}$ is the rainfall peak at the spatial resolution Δs and temporal resolution Δt. The peak attenuation ratio indicates the percentage of peak intensity lost when aggregating to a coarser resolution. Our previous study at smaller scale (Cristiano et al., 2018), highlighted that the effects of aggregation are substantial, up to −0.88 when aggregating from 100 m 1 min to 3000 m-10 min.

3.2. Rainfall Cluster Classification

To characterize the storm scale in space, dimensions of the cluster of pixels above a selected threshold are determined, as defined in Cristiano et al. (2018). The cluster dimension considered the characteristic spatial scale of the storm (Figure 2a). Cristiano et al. (2018) investigated 4 different thresholds, corresponding to the 25th, 50th, 75th, and 95th percentiles of the rain rate distribution of the selected events. The main cluster was identified at each time step. The cluster identification algorithm considers a single pixel part of a cluster if at least one of its sides is bordering the cluster. Small clusters, with an area smaller than 1% of the total catchment area, are ignored. The size of main clusters per time step is then averaged over all time steps of the rainfall event. Results presented by Cristiano et al. (2018) showed that the strongest relation with hydrological response occurred for cluster dimensions associated with the 75th percentile threshold. In particular, the hydrological response sensitivity decreases with the increase of the cluster dimension. Based on the results in Cristiano et al. (2018), the 75th percentile threshold is used in this work for Little Sugar Creek to derive cluster dimensions for the selected 28 events. The cluster dimension, defined as the cluster area above the 75th percentile threshold, characterizes the main core of the storm event that leads to runoff generation.

For temporal scale estimation, the rainfall time series for each pixel is analyzed. The "maximum wet period" above a given rainfall threshold was determined for each pixel and averaged over the total area. Like for the spatial rainfall dimension, the threshold is chosen equal to the 75th percentile of the rainfall events data set. This choice was based on the results presented by Cristiano et al. (2018), which showed a relation between cluster derived with the 75th percentile and the sensitivity of the hydrological response. The estimated value of the maximum wet period was chosen as the characteristic temporal dimension of the main storm cell of the event (Figure 2b).

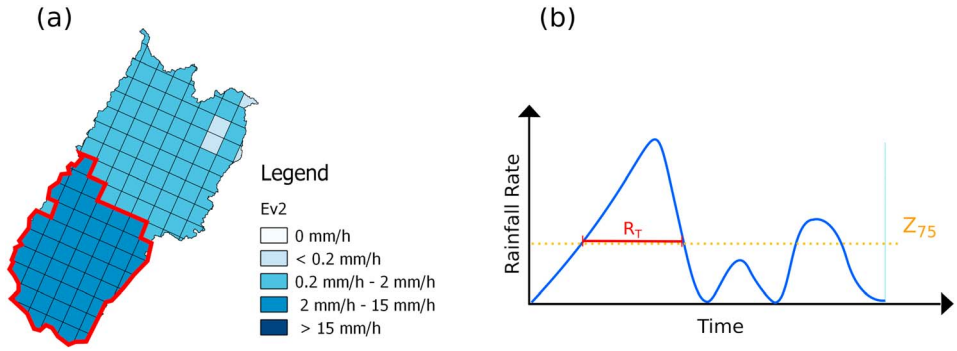

Figure 2. Representation of the methodology used to classify rainfall characteristic dimensions in space (a) and time (b).

The cluster dimension (R_s) and maximum period (R_t) above the selected threshold allow to describe the spatial and temporal variability of the storm events and were used to investigate the influence of this variability on the sensitivity of the hydrological response.

3.3. Scale Factors

3.3.1. Definition

Scale factors introduced in Cristiano et al. (2018) combine the spatial and temporal scales of rainfall (R_s and R_t) and catchment (C_s and C_t), with the spatial and temporal rainfall resolution Δs and Δt, used to drive the hydrological model. The characteristic rainfall spatial dimension R_s is represented by the rainfall cluster dimension above the 75th percentile threshold. The temporal rainfall scale R_t is characterized by the maximum period above the 75th percentile threshold. Drainage area and lag time are chosen in this study to represent the spatial and temporal catchment characteristic scales respectively (C_s and C_t).

The first scale factor α_1 (equation (2)) relates spatial rainfall (R_s) and catchment characteristic (C_s) dimensions with the spatial resolution Δs investigated. The second scale factor α_2 (equation (3)) combines spatial rainfall and temporal catchment scales, with the relative spatial and temporal resolutions. The last scale factor α_3 (equation (4)) investigates the combined effects of all four characteristics scales and rainfall resolution.

$$\alpha_1 = \frac{R_s}{\Delta s}\frac{C_s}{\Delta s} \tag{2}$$

$$\alpha_2 = \frac{R_s}{\Delta s}\frac{C_t}{\Delta t} \tag{3}$$

$$\alpha_3 = \frac{R_s C_s}{\Delta s^2}\frac{R_t C_t}{\Delta t^2} \tag{4}$$

To represent the model performance and relate it to the scale factors, the coefficient of determination R^2 was chosen as statistical indicator. the coefficient of determination R^2 is defined as the square of the Pearson correlation coefficient, which represents the linear correlation between two variables, in this case, measured and simulated model results. R^2 can vary between 0 and 1, where 1 indicates a perfect overlap between observations and simulations.

Scale factors are used to predict the level of performance, expressed in terms of coefficient of determination R^2, associated with a specific rainfall resolution, given rainfall event and catchment characteristic scales. Similarly, the scale factors can indicate the rainfall resolution, necessary to reach a certain level of performance given rainfall characteristics and basin dimensions. In order to investigate the interactions between scale factors and rainfall resolution, two values of model performance were selected as example. Values of $R^2 > 0.9$ and $R^2 > 0.8$ were chosen as representative of possible good and acceptable model performance respectively. For the two selected coefficient of determination values, scale factor thresholds, were identified for Cranbrook (Cristiano et al., 2018). These thresholds are presented in Table 3.

3.3.2. Evaluation of the Scale Factors

With the aim to evaluate the applicability of the scale factors and of their thresholds proposed by Cristiano et al. (2018; section 3.3), few indicators has been used.

Table 3
Scale Factor Thresholds Identified in Cristiano et al. (2018)

Scale Factor	$R^2 = 0.9$	$R^2 = 0.8$
α_1	100	15
α_2	40	10
α_3	3,000	450

All the output scenarios are divided four groups, depending on their interaction with the scale factor thresholds:

- True Positive (T_+), where thresholds are satisfied with good model performance and high α values;
- True Negative (T_-), which also satisfy the thresholds but with low level of performance and α values;
- False Positive (F_+), where thresholds are not satisfied because of the poor model performance in combination with high scale factor values; and
- False Negative (F_-), where low α values are not able to characterize the good model performance.

These indicators were combined in order to identify specificity, sensitivity, and positive and negative predictive values. The sensitivity S_e is defined as the ratio between True Positive scenarios and the total number of points that present good performance ($T_- + F_-$):

$$S_e = \frac{T_+}{T_+ + F_-} \tag{5}$$

Sensitivity represents cases that satisfy the thresholds with good model performance.

The specificity S_p, on the other hand, highlights the percentage of point that satisfy the scale factor threshold, among all scenarios with poor performance. Specificity is defined as follows:

$$S_p = \frac{T_-}{T_- + F_+} \tag{6}$$

High values of specificity suggest a good performance of the α scale factors, combined with a good performance of the model.

The Positive Predictive Value P_p is the percentage of values that present good performance between the values with high scale factor values:

$$P_p = \frac{T_+}{T_+ + F_+} \tag{7}$$

In the same way, the Negative Predictive Value P_n is described as the percentage of values that present poor performance between the values with low scale factor values:

$$P_n = \frac{T_-}{T_- + F_-} \tag{8}$$

High values of both positive and negative predictive value highlight a good applicability of the proposed thresholds for the α scale factors.

4. Results

4.1. Rainfall Aggregation Effect Across Range of Scales

Figure 3 shows the effect of aggregation in time and space on rainfall peaks. The peak attenuation ratio (equation (1)) is plotted against rainfall spatial and temporal resolution. The figure shows results from Cristiano et al. (2018) for the 8-km^2 catchment of Cranbrook and results derived in this work for Little Sugar Creek (111-km^2 total area). In all plots, markers indicate the median peak attenuation ratio of the selected events, while solid lines highlight the interquartile (25th percentile to 75th percentile) range and dotted lines show the minimum to maximum range.

Figure 3a presents results for Cranbrook, based on 9 rain events and 16 resolution combinations from 1 to 10 min and from 100 to 3,000 m. In this case, the effects of aggregation on the rainfall peak are very strong, leading to up to −0.8 peak reduction when aggregating in space (from 100 to 3,000 m) or in time (from 1 to 10 min). Peak reduction reaches −0.88 for spatial and temporal aggregation combined.

The effects of aggregation on the 28 events selected for Little Sugar Creek (Figure 3b) confirm that aggregation results in strong peak attenuation, reaching a maximum peak reduction of −0.6 when aggregating in

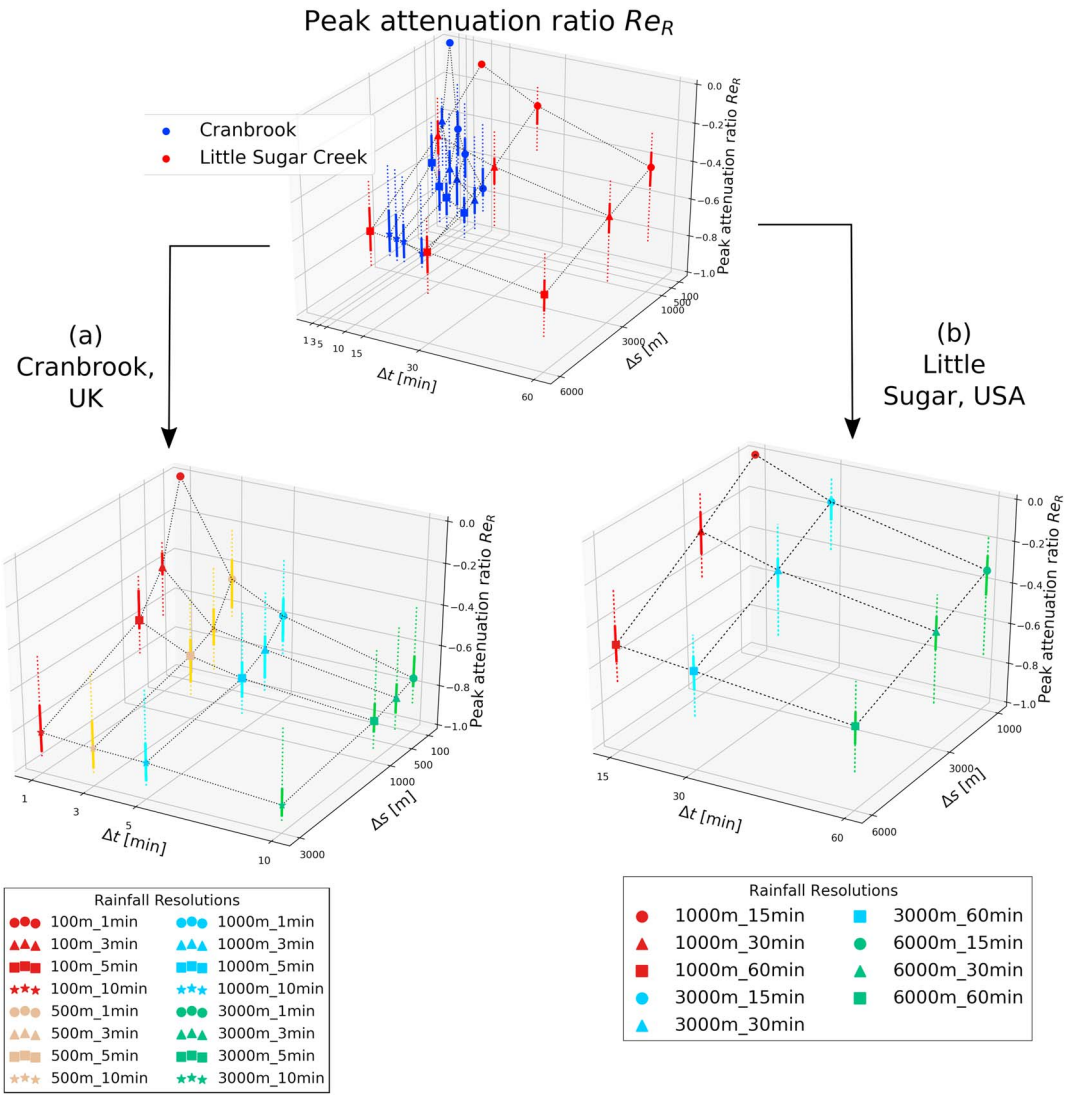

Figure 3. Peak attenuation ratios associated with rainfall aggregation. Comparison between the results presented in Cristiano et al. (2018) for Cranbrook (a) and the results obtained for Little Sugar Creek (b). Markers indicate the median, solid lines the interquartile range, and dotted lines the difference between minimum and maximum values.

space and time, from 1 km-15 min to 6 km-60 min. Spatial aggregation shows a slightly stronger effect on peak attenuation, especially at the coarser resolutions, between 3 and 6 km.

Stronger attenuation effects at smaller scale can be explained by the fact that the intermittency, defined as the alternation of dry and wet periods during the rainfall event, presents high variability at different scales. Rainfall intermittency is generally stronger for convective events (Gires et al., 2012) and it increases with finer resolution and spatial correlation increases at lower temporal resolution (Schleiss et al., 2011).

These results highlight the importance of using high-resolution spatially distributed rainfall data and suggest that using the 1 km-15 min may not be enough to reach a reasonable level of accuracy in the flow estimation.

Effects of rainfall aggregation on stream flow are also quite strong. Applying the peak attenuation ratio on the simulated flow lead to a median loss of 0.15 for subcatchment 507 and 45,022, 0.28 for 490, and 0.38 for the smallest catchment 470, when using the coarser rainfall resolution (6,000 m-60 min). Losses are less than for the rainfall, but they are still quite high and they are higher for smaller catchment, suggesting higher sensitivity for smaller basins.

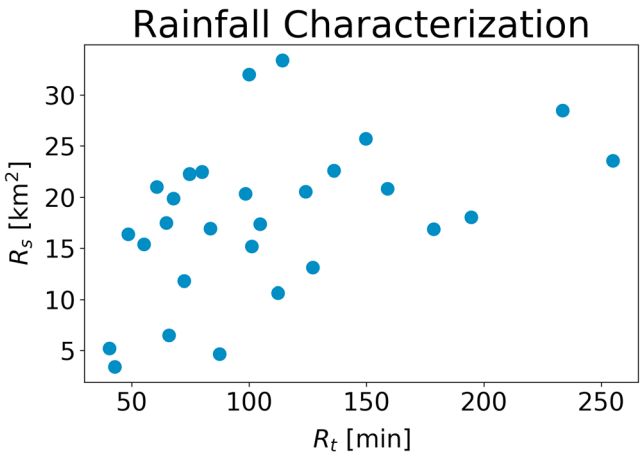

Figure 4. Spatial and temporal rainfall scales of the 28 selected events, based on cluster classification.

4.2. Rainfall Cluster Classification

The rainfall cluster classification approach described in section 3.2 was applied to the selected 28 rain events. The threshold used for cluster identification is 2 mm/hr. Figure 4 shows the spatial and temporal scales obtained by applying the cluster classification. The main cluster dimension varies from 3 to 33 km^2. Events with a short maximum wet period present a high level of intermittency, while long maximum period indicates storms that are relatively homogeneous. For the selected events, spatial scale generally increases with temporal scale, but the relation is weakly linear (Pearson correlation coefficient of 0.45). A few events were characterized by small temporal scale yet large spatial scale, corresponding to fast-moving events with a large storm core.

4.3. Scale Factors

In this section we investigate the applicability of scale factors in relation to the performance thresholds described in section 3.3, for two different situations. In the first case, the output obtained using the highest rainfall resolution, is used as a reference. This case, referred to as "Model-based analysis," characterizes the impact of rainfall input resolution on streamflow predictions. In the second case, referred as "Observation-based analysis," flow observations are used as a reference to assess model performance. This enables us to discriminate between the influence of rainfall resolution on model output versus the effect of model uncertainty on model performance. Table 4a summarizes the percentage of data points that present a good level of performance ($R^2 > 0.9$ and $R^2 > 0.8$) for both the analysis. This allows us to identify the decrease of performance due to model uncertainty. We can observe that the percentage of data points with $R^2 > 0.9$ drastically drop when we investigate the model-based analysis. This loss of accuracy is also influenced by the subcatchment characteristics and their representation in the model. Table 4b presents the loss

Table 4a
Fraction of Data Points With High Level of Performance: Aggregated Results for the Four Subcatchments

	Model-based		Observation-based	
Resolution	$R^2 > 0.9$	$R^2 > 0.8$	$R^2 > 0.9$	$R^2 > 0.8$
1000 m-15 min	1.00	1.00	0.38	0.55
1000 m-30 min	0.79	0.89	0.21	0.47
1000 m-60 min	0.20	0.55	0.05	0.21
3000 m-15 min	0.98	0.99	0.36	0.54
3000 m-30 min	0.79	0.89	0.27	0.48
3000 m-60 min	0.23	0.55	0.06	0.24
6000 m-15 min	0.91	0.96	0.25	0.53
6000 m-30 min	0.73	0.87	0.24	0.46
6000 m-60 min	0.27	0.60	0.08	0.28

Table 4b
Fraction of Data Points With High Level of Performance: The Resolution 1,000 m-30 min Is Presented as Example, to Demonstrate Differences Between Subcatchments

Resolution 1,000 m-30 min				
	Model-based		Observation-based	
Subcatchment ID	$R^2 > 0.9$	$R^2 > 0.8$	$R^2 > 0.9$	$R^2 > 0.8$
409	0.89	1.00	0.29	0.64
45022	1.00	1.00	0.36	0.64
470	0.29	0.57	0.04	0.04
507	1.00	1.00	0.18	0.57

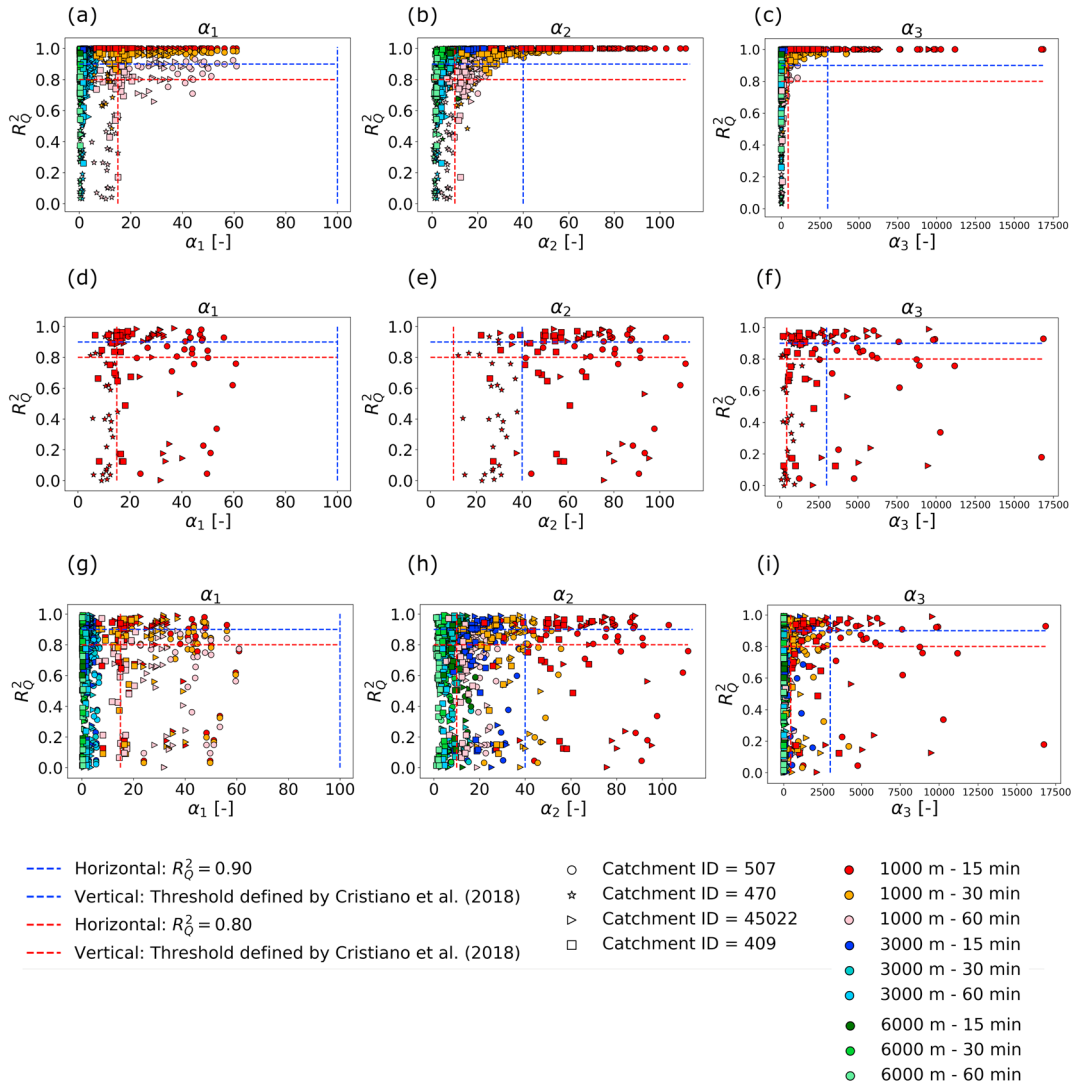

Figure 5. Coefficient of determination R^2 as function of scale factors. Colors indicate different rainfall resolutions and symbols represent different subcatchments. (a–c) Model-based analysis, (d–f) observation-based analysis focusing on the highest rainfall resolution investigated, and (g–i) observation-based analysis for all rainfall resolutions.

Table 5a
Indicators Estimated for Model-Based Analysis

	Model-based analysis					
	$R^2 > 0.9$			$R^2 > 0.8$		
Indicators	α_1	α_2	α_3	α_1	α_2	α_3
Sensitivity S_e	0.00	0.00	0.00	0.00	0.16	0.01
Specificity S_p	1.00	1.00	1.00	1.00	1.00	1.00
Positive Predictive Value P_p	0	1.00	1.00	0.00	1.00	1.00
Negative Predictive Value P_n	0.50	0.50	0.50	0.29	0.33	0.30

of accuracy for the resolution 1000 m - 30 min as example to illustrate the influence of different subcatchments. The model is able to represent better sub catchment 45022, which shows high level of performance, while the small basin of Little Hope (470) is not well represented in the model, as indicated by a low level of performance, for both model-based and observation-based analysis.

4.3.1. Model-Based Analysis

Figures 5a–5c show results for the model-based analysis, where model performance in terms of R^2 is plotted as a function of the three dimensionless scale factors. In this case, R^2 represents the correlation between the hydrograph obtained using the highest-resolution rainfall data available and the hydrograph derived from the aggregated rainfall data. Blue and red dotted lines highlight the α thresholds described in section 3.3 (corresponding respectively to $R^2 = 0.9$ and $R^2 = 0.8$), colors indicate different resolution combinations and markers represent different subcatchments. Low α values are associated with low rainfall resolution relative to rainfall and catchment scale. As expected, model performance generally improves for higher α values. Still, a large number of results are concentrated in the upper left quadrant; for those events, model performance appears independent of rainfall input resolution.

The plots show that of the three scale factors, α_2 is best capable of separating good performance versus low-performance events. The thresholds that were identified in the previous study, based on a range of higher rainfall input resolutions, seem to apply reasonably well for the case of Charlotte: especially for α_2 and α_3, the R^2=0.9 performance threshold correctly identifies events with performance above and below the α threshold value. The threshold value for R^2=0.8 performance seems too strict for the Charlotte case: it could be relaxed to higher values of 20 and 500 for α_2 and α_3, respectively. The performance indicators described in section 3.3 are presented in Table 5a. Performance scores are generally better for α_2 than for α_1 and α_3, showing that α_2 is better able to predict good model performance. Negative predictive values generally indicates a good performance, and it identifies low performance associated with insufficient rainfall resolution. Highest sensitivity, 0.16, is reached for α_2 at the R^2 >0.8 performance threshold, which implies that α thresholds are generally strict and reject many events with good performance. Performance for those events is relatively insensitive to rainfall input resolution. In section 4.4, we will investigate in more detail what explains low sensitivity for these events.

4.3.2. Observation-based Analysis

Figures 5d–5f show model performance comparing model results using the radar rainfall as input to flow observations. Model performance is generally good for catchment 45022 (21 out of 28 events present $R^2 > 0.8$) and generally poor for the smallest subcatchment (6 out of 28 events present $R^2 > 0.8$). Performance for the other subcatchments falls within this range. For the Little Hope (470) subcatchment, low model

Table 5b
Indicators Estimated for Observation-Based Analysis

	Observation-based analysis					
	$R^2 > 0.9$			$R^2 > 0.8$		
Indicators	α_1	α_2	α_3	α_1	α_2	α_3
Sensitivity S_e	0.00	0.13	0.07	0.21	0.53	0.25
Specificity S_p	1.00	0.92	0.96	0.76	0.55	0.82
Positive Predictive Value P_p	0	0.36	0.37	0.48	0.55	0.59
Negative Predictive Value P_n	0.73	0.74	0.74	0.48	0.52	0.51

performance is probably associated with the small scale of the catchment: model and rainfall resolutions are too low to properly represent the catchment response. Moreover, the insufficient representation of storm drain for this subcatchment leads to a poor model performance, as already highlighted in Wright et al. (2014a). For this reason, Little Hope is excluded from the observation-based analysis. For α_2 and α_3 about 30% of the data points above the α thresholds show a low level of performance, mostly for catchments 509 and 409. This implies the model is unable to properly reproduce hydrological response for these events for reasons other than rainfall resolution. This will be investigated more in detail in section 4.4.

In Figures 5g–5i, R^2 values are plotted for all rainfall resolutions, in relation to the scale factors and their threshold values. Performance is generally poorer than in the model-based analysis, as would be expected, since model uncertainty is now added to the effect of varying rainfall input resolution. The number of events that do not reach indicated performance levels is higher. The plots show that α values are less capable of separating good and poor performance, with 9% and 13% of events above the threshold values for α_2 and α_3, respectively, associated with R^2=0.8 not meeting the required performance. Comparing indicators of model- and measurement-based analysis (Tables 5a and 5b), we observe a general increase of sensitivity and of negative predictive value and a decrease of specificity and positive predictive value. This behavior can be explained by errors introduced by the model. The scale factor α_2 is still the one that better represents the influence of rainfall and catchment scale on the hydrological response. This is confirmed by the high values of sensitivity. The poor performance of α_1 to Little Sugar Creek study case highlights the importance of temporal dimensions, for both rainfall and catchment classification.

4.4. Role of Rainfall Characteristics

The decrease in model performance associated with rainfall resolution is relatively small compared to the difference in performance between events and catchments, presented in Figures 5g–5i. Low model performance for catchment 407 is clearly associated with insufficient rainfall resolution for a basin of this small scale. For the other catchments, performance is good ($R^2 > 0.8$) for 70% of events but falls as low as zero for some events.

In this section we investigate properties of rain events associated with low performance in order to identify which characteristics could explain this poor performance. In particular, we focus on those events that present high values of α_2 and low values of R^2. Figure 6 shows α_2, which was previously shown to be the most reliable scale factor, in relation to the coefficient of determination. Different colors highlight the influence of different rainfall characteristics. On the left side (Figures 6a–6d) are plotted only the high rainfall resolution (1000 m-15 min) results for the three larger subcatchments. On the right side (Figures 6e–6h) all the resolutions are presented. The following characteristics are investigated: maximum wet period above the 75th percentile threshold R_t (Figures 6a and 6e), total rainfall depth in millimeters (Figures 6b and 6f), ratio between the maximum wet period above the 75th percentile threshold and the total duration of the event $\%T$ (Figures 6c and 6g) and the maximum 15-min intensity in a single pixel, expressed in millimeters per hour (Figures 6d and 6h).

Rainfall events with a short maximum wet period (darker color points in Figures 6a and 6e) generally present lower performance, while lighter points (long maximum wet period) show a higher coefficient of determination. Sixty-three percent of the data points with low performance present short maximum wet periods ranging between 40 and 112 min.

Total rainfall depth can explain the low level of model performance for some rainfall events. Low values of total rainfall depth lead to a low level of performance: for the highest rainfall resolution, 80% of the rainfall events with a $R^2 < 0.8$ present a total rainfall depth between 8 and 37 mm. The percentage of data point drops to 72%, when considering all the rainfall resolutions. Low performance for small rainfall events and events with short-lasting storm cores could be explained by insufficient capability of the model to represent runoff generation processes (initial wetting losses, infiltration, and runoff velocities) or initial wetting conditions of the catchment. We investigated the latter by running not only the events separately but also the full 15-year time series. This resulted in only minor changes in the modeled hydrographs. Hence, insufficient representation of runoff processes seems the most likely explanation for poor performance associated with this type of events.

Figures 6c and 6g show that also $\%T$, the ratio between the maximum wet period and the total duration of the event, seems to affect the level of performance of the model. This parameter represents the intermittency

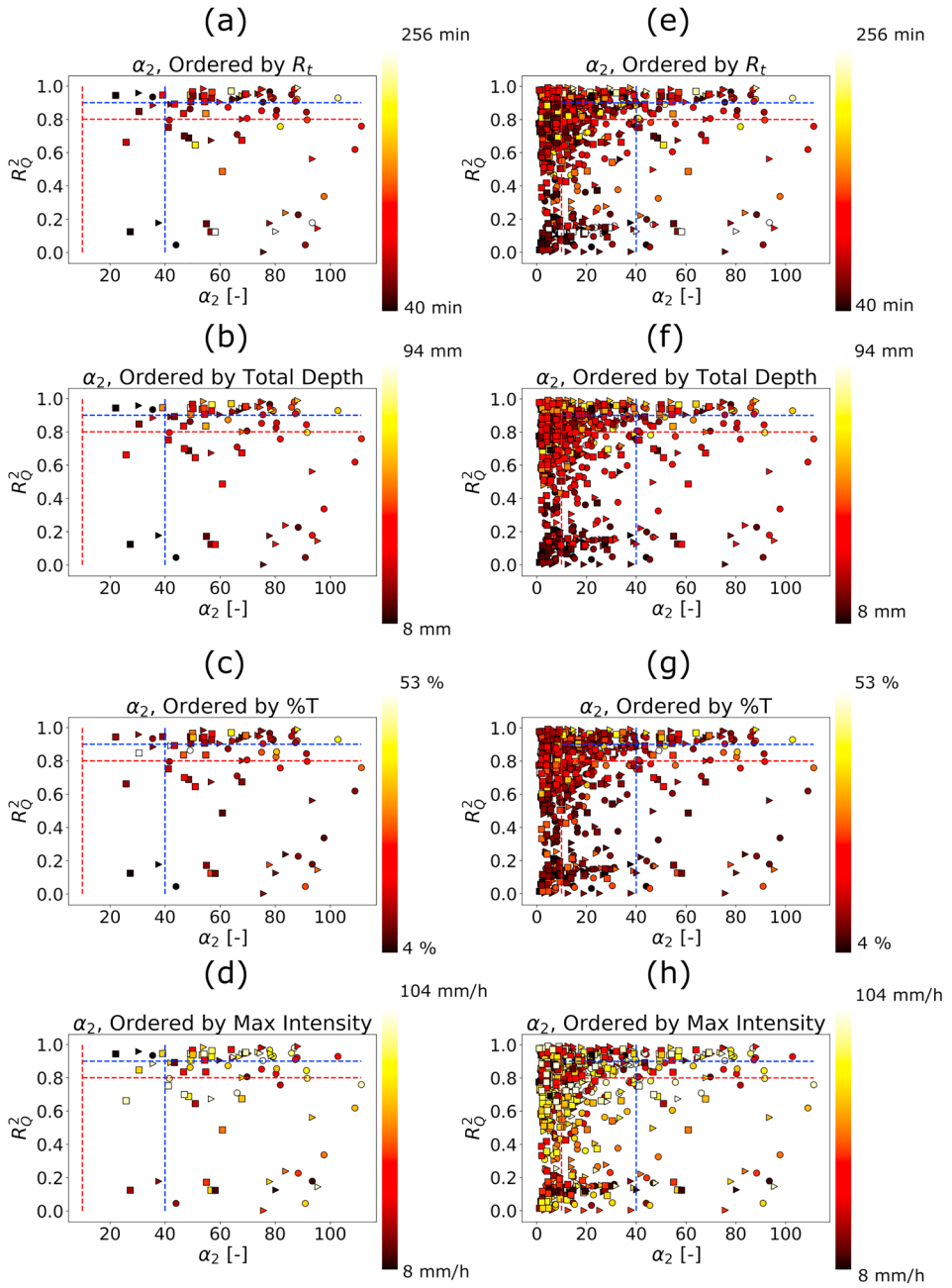

Figure 6. Influence of rainfall event characteristics on the hydrological response sensitivity. Different colors indicated different values maximum wet period above the 75th percentile, total rainfall depth, percentage of wet period over the total duration, and maximum intensity for the highest resolution 1,000 m-15 min (a–d) and for all the resolution investigated (e–h).

of the rainfall event, estimating for how long the rainfall pixel is above the threshold, in relation to the total duration of the event. Most of the low-performance data points (73%) present low values of *%T* (in a range between 4% and 20%), showing that intermittent events in particular are poorly captured by the model, probably due to insufficient representation of runoff generating processes.

The last rainfall characteristic investigated is the maximum pixel intensity. In this case, many of the data points with low model performance (43%) present high values of maximum rainfall intensity, but the effect is less prominent compared to intermittency and rainfall volume.

The rainfall characteristics investigated in this section can partially describe the coexistence of poor model performance and high scale factor values. The combination of low intermittency, low total depth, short maximum wet period, and high maximum intensity describe the rainfall events that lead to poor performance, dominated by shortcomings of the model, in particular runoff generation processes, over the effects of rainfall resolution.

5. Summary and Conclusions

In this study we investigated the relationships between rainfall and catchment scales and the sensitivity of urban hydrological response to rainfall data resolution. We used cluster classification and scale factors proposed by Cristiano et al. (2018) for a different range of spatial and temporal scales and in a watershed with very different land use and hydroclimatological characteristics. Specifically, Cristiano et al. (2018) examined a small (8 km^2) heavily urbanized catchment in the United Kingdom, affected primarily by short and intense convective storms, this study examines Little Sugar Creek (111 km^2) in Charlotte, North Carolina (USA). Little Sugar Creek typifies more recent storm water infrastructure and is subject to widely varied storms, chiefly tropical cyclones and summertime organized thunderstorm systems. These differences mean that Little Sugar Creek thus constitutes a useful addition second case study for this methodology, since it allows us to evaluate the generalizability of findings to new locations and scales. Scale factor values support prediction of low model performance associated with insufficient rainfall input resolution support and, conversely, identification of critical rainfall resolutions required for high model performance. Little Sugar Creek offers the unique opportunity to work with a physically based distributed hydrological model and 15 years of radar rainfall data and local streamflow observations. This enables discriminating between uncertainties in the model representation versus effects of changes in rainfall input, in an observation-based analysis. Twenty-eight rainfall events were selected as input for the model, with the aim to analyze the applicability of scale factors developed in an earlier study and to identify critical rainfall scales, in a different physiographic and climatic setting, over a different range of scales and much larger set of rainfall conditions.

Cluster classification and maximum wet period above a selected threshold were used to define characteristic rainfall spatial and temporal scales. This approach enables a simple and fast estimation of the core of rainfall events, which strongly affects the hydrological response. The proposed scale factors are based on dimensionless ratios between storm and catchment scales versus rainfall space-time resolutions. Previous studies aiming to identify critical rainfall resolution for hydrological response modeling have focused on catchment, storm, and model scales individually (Berne et al., 2004; Gires et al., 2012; Ochoa-Rodriguez et al., 2015; Rafieeinasab et al., 2015). Scale factors proposed in this work combine ratios of rainfall and catchment scales, increasing predictive power for relations between rainfall resolution and model performance compared to previous urban hydrologic modeling studies. The scale factors that combine rainfall and catchment scales showed higher specificity than the scale factor based on the spatial rainfall and catchment scale only. These results confirm the crucial importance of temporal scale for both rainfall and catchment and, in particular, it highlights that storm events with strong rainfall intermittency were poorly represented by the scale factors. It also suggests that improving the temporal rainfall characterization, including a description of the rainfall event intermittency could improve the performance of the scale factors.

Scale factor thresholds, derived for Cranbrook to identify the required rainfall resolution, were tested for Little Sugar Creek, showing a good applicability to different catchment scales and different geomorphological and climatological areas, especially in the model-based analysis. Rainfall resolution of 1,000 m-15 min clearly showed to not be sufficient for the smallest subcatchment of Little Hope. In this case, due to the small catchment scale, a higher rainfall resolution is required. This analysis highlights the contribution of model uncertainty compared to rainfall resolution to overall model performance. For $R^2 = 0.9$, thresholds correctly identify events with high performance. In the observation-based analysis, were observed some cases with low performance for small rainfall events. A possible explanation could be the insufficient representation of runoff generation processes and initial wetting conditions in the model. This could also explain why intermittent events are generally poorly captured by the model.

Although results have shown a good applicability of scale factors to a large urban catchment, more study cases and levels of model complexity, in particular in relation runoff generating processes, need to be investigated to fully generalize the developed approach. Additional locations, with local flow measurements and high-resolution rainfall radar data, should be used to generalize the new methodology and obtain a tool to

identify the minimum required rainfall resolution for a specific catchment and to define the reliability of model results.

Acknowledgments
The authors would like to thank Mary Lynn Baeck for making available the NEXRAD radar rainfall data sets used in this study. The first author is grateful to Leon Riesebos for the technical support in running the simulations. This work has been supported by the EU INTERREG IVB through funding of the RainGain Project (http://www.raingain.eu).

References

Berne, A., Delrieu, G., Creutin, G. D., & Obled, C. (2004). Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology*, *299*, 166–179.

Bruni, G., Reinoso, R., van de Giesen, N. C., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2015). On the sensitivity of urban hydrodynamic modelling to rainfall spatial and temporal resolution. *Hydrology and Earth System Sciences*, *19*, 691–709.

Cristiano, E., ten Veldhuis, M.-C., Gaitan, S., Ochoa Rodriguez, S., & van de Giesen, N. (2018). Critical scales to explain urban hydrological response: An application in Cranbrook, London. *Hydrology and Earth System Sciences*, *22*(4), 2425–2447.

Cristiano, E., ten Veldhuis, M.-C., & van de Giesen, N. (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.

Downer, C. W., & Ogden, F. L. (2003). Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment. *Water Resources Research*, *39*(3), 1045. https://doi.org/10.1029/2002WR001439

Downer, C. W., & Ogden, F. L. (2004). Gssha: Model to simulate diverse stream flow producing processes. *Journal of Hydrologic Engineering*, *9*(3), 161–174.

Driscoll, E. D., Palhegyi, G. E., Strecker, E. W., & Shelley, P. E. (1989). Analysis of storm event characteristics for selected rainfall gages thoughout the United States (p. 48). Washington DC: U.S. Environmental Protection Agency.

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N. E., Quirmbach, M., Vaes, G., & Vieux, B. (2004). Towards a roadmap for use of radar rainfall data use in urban drainage. *Journal of Hydrology*, *299*, 186–202.

Faures, J. M., Goodrich, D. C., Woolhiser, D. A., & Sorooshian, S. (1995). Impact of small scale spatial rainfall variability on runoff modelling. *Journal of Hydrology*, *173*, 309–326.

Ferraris, L., Gabellani, S., Rebora, N., & Provenzale, A. (2003). A comparison of stochastic models for spatial rainfall downscaling. *Water Research Resources*, *39*(12), 1368. https://doi.org/10.1029/2003WR002504

Gires, A., Onof, C., Maksimovic, C., Schertzer, D., Tchiguirinskaia, I., & Simoes, N. (2012). Quantifying the impact of small scale unmeasured rainfall variability on urban hydrology through multifractal downscaling: A case study. *Journal of Hydrology*, *442-443*, 117–128.

Gironás, J., Niemann, J., Roesner, L., Rodriguez, F., & Andrieu, H. (2010). Evaluation of methods for representing urban terrain in storm-water modeling. *Journal of Hydrologic Engineering*, *15*, 1–14.

Krajewski, W. F., Kruger, A., Lawrence, R., Smith, J. A., Bradley, A. A., Steiner, M., et al. (2007). Towards better utilization of NEXRAD data in hydrology: An overview of hydro-NEXRAD. World Environmental and Water Resources Congress.

Leijnse, H., Uijlenhoet, R., & Stricker, J. N. M. (2007). Rainfall measurement using radio links from cellular communication networks. *Water Research Resources*, *43*, W03201. https://doi.org/10.1029/2006WR005631

Leijnse, H., Uijlenhoet, R., van de Beek, C. Z., Overeem, A., Otto, T., Unal, C. M. H., et al. (2010). Precipitation measurement at Cesar, The Netherlands. *Journal of Hydrometeorology*, *11*, 1322–1329.

Muthusamy, M., Schellart, A., Tait, S., & Heuvelink, G. B. M. (2017). Geostatistical upscaling of rain gauge data to support uncertainty analysis of lumped urban hydrological models. *Hydrology and Earth System Sciences*, *21*, 1077–1091. https://doi.org/10.5194/hess-21-1077-2017

Niemczynowicz, J. (1988). The rainfall movement—A valuable complement to short-term rainfall data. *Journal of Hydrometeorology.*, *104*, 311–326.

Ochoa-Rodriguez, S., Wang, L. P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., et al. (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *Journal of Hydrology*, *531*(2), 389–407.

Pina, R., Ochoa-Rodriguez, S., Simones, A., Sa Marques, A., & Maksimovik, C. (2014). Semi-distributed or fully distributed rainfall-runoff models for urban pluvial flood modelling? In *13th International Conference on Urban Drainage, Sarawak, Malaysia, 7-12 September 2014* (pp. 2073–4441).

Pina, R., Ochoa-Rodriguez, S., Simones, A., Sa Marques, A., & Maksimovik, C. (2016). Semi- vs fully- distributed urban stormwater models: Model set up and comparison with two real case studies. *Water*, *8*, 58.

Rafieeinasab, A., Norouzi, A., Kim, H., Nazari, B., Seo, D., Lee, H., et al. (2015). Toward high-resolution flash flood prediction in large urban areas analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling. *Journal of Hydrology*, *531*(2), 370–388.

Restrepo-Posada, P. J., & Eagleson, P. S. (1982). Identification of independent rainstorms. *Journal of Hydrology*, *55*(1), 303–319.

Salvadore, E., Bronders, J., & Batelaan, O. (2015). Hydrological modelling of urbanized catchments: A review and future directions. *Journal of Hydrology*, *529*(1), 61–81.

Schilling, W. (1991). Rainfall data for urban hydrology: What do we need? *Atmospheric Research*, *27*, 5–21.

Schleiss, M., Jaffrain, J., & Berne, A. (2011). Statistical analysis of rainfall intermittency at small spatial and temporal scales. *Geophysical Research Letters*, *38*, L18403. https://doi.org/10.1029/2011GL049000

Sempere-Torres, D., Corral, C., Raso, J., & Malgrat, P. (1999). Use of weather radar for combined sewer overflows monitoring and control. *Journal of Environmental Engineering*, *125*(4), 372–380.

Seo, B.-C., Krajewski, W. F., Kruger, A., Domaszczynski, P., Smith, J. A., & Steiner, M. (2010). Radar-rainfall estimation algorithms of hydro-NEXRAD. *Journal of Hydroinformatics*, *13*(2), 277.

Smith, A. J., Baeck, M. L., Morrison, J. E., Sturevant-Rees, P., Turner-Gillespie, D. F., & Bates, P. D. (2002). The regional hydrology of extreme floods in an urbanizing drainage basin. *American Meterological Society*, *3*, 267–282.

Smith, A. J., Baeck, M. L., Villarini, G., Welty, C., Miller, A. J., & Krajewski, W. F. (2012). Analyses of a long term, high resolution radar rainfall data set for the baltimore metropolitan region. *Water Resources Research*, *48*, W04504. https://doi.org/10.1029/2011WR010641

Smith, B. K., Smith, A. J., Baeck, M. L., Villarini, G., & Wright, D. B. (2013). Spectrum of storm event hydrologic response in urban watersheds. *Water Resources Research*, *49*, 2649–2663. https://doi.org/10.1002/wrcr.20223

ten Veldhuis, M.-C., & Schleiss, M. (2017). Statistical analysis of hydrological response in urbanising catchments based on adaptive sampling using inter-amount times. *Hydrology and Earth System Sciences*, *21*(4), 1991–2013.

ten Veldhuis, M.-C., Zhou, Z., Yang, L., Liu, S., & Smith, J. (2018). The role of storm scale, position and movement in controlling urban flood response. *Hydrology and Earth System Sciences*, *22*(1), 417–436.

Thorndahl, S., Einfalt, T., Willems, P., Nielsen, J. E., ten Veldhuis, M.-C., Arnbjerg-Nielsen, K., et al. (2017). Weather radar rainfall data in urban hydrology. *Hydrology and Earth System Sciences*, *21*(3), 1359–1380.

Tokarczyk, P., Leitao, J. P., Rieckermann, J., Schindler, K., & Blumensaat, F. (2015). High-quality observation of surface imperviousness for urban runoff modelling using UAV imagery. *Hydrology and Earth System Sciences*, *19*, 4215–4228.

van de Beek, C. Z., Leijnse, H., Stricker, J. N. M., Uijlenhoet, R., & Russchenberg, H. W. J. (2010). Performance of high-resolution x-band radar for rainfall measurement in the Netherlands. *Hydrology and Earth System Sciences*, *14*, 205–221.

Wright, D. B., Smith, J. A., & Baeck, M. L. (2014a). Flood frequency analysis using radar rainfall fields and stochastic storm transposition. *Water Resources Research*, *50*, 1592–1615. https://doi.org/10.1002/2013WR014224

Wright, D. B., Smith, J. A., Villarini, G., & Baeck, M. L. (2013). Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. *Journal of Hydrology*, *488*, 150–165.

Wright, D. B., Smith, J. A., Villarini, G., & Baeck, M. L. (2014b). Long-term high-resolution radar rainfall fields for urban hydrology. *JAWRA Journal of the American Water Resources Association*, *50*(3), 713–734. https://doi.org/10.1111/jawr.12139

Xuereb, k., & Green, J. (2012). Defining independence of rainfall events with a partial duration series approach. In *Proceedings of the 34th Hydrology and Water Resources Symposium, Engineers Australia* (pp. 169–176).

Yang, L., Smith, J. A., Baeck, M. L., & Zhang, Y. (2016). Flash flooding in small urban watersheds: Storm event hydrological response. *Water Resources Research*, *52*, 4571–4589. https://doi.org/10.1002/2015WR018326

Zhou, Z., Smith, J. A., Yang, L., Baeck, M. L., Chaney, M., Ten Veldhuis, M.-C., et al. (2017). The complexities of urban flood response: Flood frequency analyses for the charlotte metropolitan region. *Water Resources Research*, *53*, 7401–7425. https://doi.org/10.1002/2016WR019997

Zoppou, C. (2000). Review of urban storm water models. *Environmental Modelling & Software*, *16*, 195–231.