Author response: “Uncertainty in Hydrological Signatures” by I. K. Westerberg and H. K. McMillan

We thank all referees for their constructive comments that helped to improve our paper. The referee comments are given in blue text below with our responses in black. Changes made to the original manuscript are detailed in italics in the responses, and in the track-changes version of the manuscript included in the end of this document.

Response to Referee 1, A. Viglione

The paper shows how uncertainties in catchment rainfall and runoff measurement, interpolation and extrapolation propagate into uncertainties in hydrologic signatures, which are widely used in hydrology. The Authors use a methodology for uncertainty propagation based on Monte Carlo simulations. They consider sources of uncertainty and uncertainty models proposed in the literature. The methodology is applied to two small catchments in England and New Zealand.

I really liked reading this paper, which is well written and inspiring. I definitively see the need for more publications of this kind in order to build up a more generalised understanding of the uncertainties in the data (and in how the data are used) which are at the basis of hydrologic studies. I am therefore supportive for the publication of the paper in HESS.

Response: We thank Alberto Viglione for his helpful review and his positive comments about our paper.

I just have a couple of suggestions which may help to improve the paper and which require little additional work for the Authors:

- The amount of information provided in the paper is a lot and it would be nice to summarise it in tables. As in Table 1 the signatures are explained, it would be nice to have a table that lists all sources of uncertainty considered and the references in the literature where they have been discussed. More importantly, it would be nice to have a final table that summarises the major (dominant) sources of uncertainty for each signature as well as the relative uncertainty ranges found for the two catchments under study.

Response: Thank you for this suggestion, we agree that it is a good idea to have an overview of the uncertainty sources with references to the literature. We have included a new table (Table 1, referred to in the first paragraph in Section 3.1 in the revised manuscript,) that shows the sources of uncertainty we considered in this paper, the methods used to estimate them and literature references for the methods where applicable.

Table 1. Sources of uncertainty considered in this study, and the methods used for estimation.

Variable/Signature	Uncertainty component	Estimation method	Reference if applicable
Rainfall	Point uncertainty	Normal distribution with σ a function of rain rate.	Ciach (2003)
	Interpolation uncertainty	Subsampling from a dense network of rain	

		gauges	
	Equipment malfunction	Rainfall data with/without QC	Wood et al. (2000)
Flow	Discharge uncertainty in gaugings	Analysis of stations with stable ratings	Coxon et al. (2015)
	Stage uncertainty in gaugings	Uniform distribution ±5 mm	McMillan et al. (2012)
	Rating-curve uncertainty	Voting Point likelihood method	McMillan and Westerberg (2015)
Recession analysis	Flow data time step	Tested hourly vs daily	
	Seasonality of response	Tested using all data or split by season	Shaw and Riha (2012)
Rainfall-runoff threshold	Effects of baseflow	Tested with/without baseflow separation	Gustard et al. (1992)
	Rainfall event definition	Tested with/without inclusion of smaller events	

We have also included a final table (Table 3 in the revised manuscript) at the end of the results section that summaries the dominant uncertainty sources and the uncertainty magnitudes and characteristics (see also next response below). We included a new subsection (4.3) to describe the table and summarise the findings, see below.

4.3 Summary of the signature uncertainties

To summarise our results, we tabulated examples of each signature type together with their dominant uncertainty sources and summary statistics of the total uncertainty distribution, for each catchment (Table 3). Our aim is to allow easy comparison of the signature uncertainties in our study, with those of other studies. We therefore chose commonly used distribution statistics, i.e. the first three distribution moments (mean, standard deviation, skewness); and the half-width of the 5–95 percentile range, which is commonly quoted in uncertainty studies (e.g. McMillan et al, 2012). We hope that authors of future studies will consider using similar statistics, to enable the community to compile a generalised understanding of signature uncertainties across different catchments, scales and landscapes.

Table 3 Dominant uncertainty sources and uncertainty characteristics

Signature type		**Catchment[1]**	**Dominant uncertainty source**	**Uncertainty characteristics**			
				Half-width of 5–95 percentile range (%)	Mean (=bias) (%)	Std. (%)	Skew-ness (-)
Flow distribution	Average flow conditions (Q_{MEAN})	M	Rating-curve uncertainty	11.1	-0.4	6.8	0.32
		B	Rating-curve uncertainty	12.7	-2.4	7.7	-0.03
	Low flow percentiles (Q_{95})	M	Discharge gauging uncertainty	23.8	-1.2	14.6	0.47
		B	Rating-curve uncertainty	39.5	-1.1	23.8	0.45
	High flow percentiles ($Q_{0.1}$)	M	Rating-curve uncertainty	22.8	-8.3	16.6	1.54
		B	Rating-curve uncertainty	19.6	0.0	12.0	0.13

Events	Event frequency and duration (Q_{HD})	M	Threshold value, which depends on rating-curve uncertainty	6.9	2.3	3.3	1.30
		B	Threshold value, which depends on rating-curve uncertainty	21.6	-5.1	13.1	0.57
Flow dynamics	Base Flow Index (BFI)	M	Rating-curve uncertainty	11.6	3.4	7.1	-0.11
		B	Rating-curve uncertainty	8.5	-2.3	5.1	-0.19
	Slope of Flow Duration Curve (S_{FDC})	M	Rating-curve breakpoint location	28.8	16.9	17.4	0.46
		B	Rating-curve uncertainty	6.0	-3.2	3.7	-0.18
	Variability of extreme flows (Q_{HV})	M	Rating-curve uncertainty	41.9	-1.0	30.4	2.30
		B	Rating-curve uncertainty	37.0	6.5	23.0	0.75
	Recession analysis (b hourly)	M	Calculation time step	9.9	-3.1	6.3	0.38
		B	Rating-curve uncertainty	14.9	5.1	8.9	0.72
Rainfall-runoff	Total runoff ratio (RR^2)	M	Rating-curve uncertainty	14.6	-0.3	9.0	0.26
		B	Rating-curve uncertainty	13.3	-2.0	8.1	0.02
	Rainfall-runoff threshold (threshold location[3])	M	Rainfall interpolation uncertainty	17.3	16.3	17.2	5.88
		B	-	-	-	-	-
Rainfall	Mean annual precipitation ($P_{MA}{}^2$)	M	Interpolation uncertainty	10.0	0.3	5.7	0.22
		B	Interpolation uncertainty. (Equipment malfunction)	4.6	-0.4	2.7	0.34
	Standard deviation of precipitation ($P_{STD}{}^2$)	M	Interpolation uncertainty	8.0	9.5	4.4	1.55
		B	Interpolation uncertainty. (Equipment malfunction)	4.9	4.6	2.9	0.67

[1] M = Mahurangi, B = Brue

[2] These signatures were calculated using 1 gauge/45 km^2 and including point error

[3] This signature was calculated for the total uncertainty scenario in Fig. 10

- Reading the title of the paper I would have expected more discussion on generalisation of results. I was involved in editing a book on runoff prediction in ungauged basins (Blöschl et al., 2013, already cited in the paper), where an assessment of uncertainty of regionalisation methods was attempted based on a literature review of many studies around the world. Let’s assume that in the next years many researches will perform similar studies on uncertainty in hydrologic signatures and that the Authors will be asked to synthetise these works (and try to understand the effect of climate, catchment scale, dominant hydrologic processes, antropogenic influence, etc...). What information would the Authors like to find in these papers? How this information should be organised and presented? This may be discussed in the conclusion and the final table referred to in the previous point could be an example of what the Authors would like to find in other papers on the subject. In other words, I believe that this paper could aim at setting a standard for studies on uncertainty in hydrologic signatures.

Response: Thank you for this good suggestion. Observational uncertainties are in general highly dependent on local site conditions and measurement methods, therefore it is important to include such information (measurement equipment, metadata about station characteristics, out-of-bank levels, temporal changes in site characteristics, etc.) together with information about catchment size, scale, human impacts, dominant hydrologic processes, etc (as suggested). The place-specific nature of the uncertainties will likely impede the possibilities to draw general conclusions about some influencing factors; however, we believe that such generalisation attempts are important and that valuable insights could be gained in future review studies.

Regarding information about the magnitudes of the uncertainties we believe that is important to not only report information about upper/lower uncertainty bounds, but also information about the shape of the estimated uncertainty distributions, e.g. by using histograms or boxplots as in this study. Therefore we have calculated four summary statistics (mean, std, skewness and halfwidth of the 5-95 percentile range) describing the uncertainty distributions and reported them together with the dominant uncertainty sources in the new summary results table described in the previous response (Table 3 in the revised manuscript). We reported the uncertainty magnitudes for one representative signature per category and catchment, to keep the table compact in terms of summarising the findings. We have noted that other papers could follow this example in reporting signature uncertainties, recommending that statistics describing the shape of the uncertainty distributions are always reported (see new text and table in the previous response).

Minor comments:

Page 4237, lines 21 and 24: I get confused here. “The main aim of this paper was...” refers to Juston et al. (2014) while “The objectives of this paper were:” refers to the present paper. Am I right? Maybe a rewording could help the reader here.

Response: The first part about the main aim of our paper we included to say up front that the results will be sensitive to the uncertainty estimation technique and the understanding of the uncertainty sources (also discussed by Juston et al). We have reworded the first part to clarify this:

This paper was focused on signature uncertainty rather than data uncertainty; we stress that alternative data uncertainty assessment methods could be used where the perceptual understanding of the uncertainty sources is different.

We have also introduced a new paragraph in the discussion in section 5.2 about the sensitivity of the results relating to the uncertainty estimation methods, as also suggested by the other referees (see response to Referee 2, H. Gupta below).

Page 4238 line 15 and page 4239 line 6: why missing precipitation values have been infilled with two different procedures in the two catchments? I guess the reason is because that was done in previous works but the text doesn’t state it. Moreover, methods of infilling rainfall data are not considered in the uncertainty analysis, why?

Response: Yes, in the New Zealand catchment they had already been infilled in a previous project, and we have noted this in the text (Section 2.1).

Missing rainfall values were available from a previous study that had infilled them using linear correlation with a nearby site.

Methods for infilling rainfall data could also have been considered, although we believe this would have had a small impact on the results since there were small differences between the different methods we tried.

Page 4251, lines 9-12: I do not understand why events defined using a threshold related to the mean or median flow are more sensitive to rating curve uncertainty than events defined using a flow percentile threshold. What percentile is preferable? In the end the median is also a percentile, why isn't it good?

Response: It is when the threshold is defined using a *multiplier* of the median value, instead of the actual flow percentile (or median) value itself, that it becomes sensitive to the rating-curve uncertainty. This is because the (uncertain) gradient of the rating curve greatly impacts on the flow percentile equivalent to the threshold value. We explained this in better wording in the results section on P4247 Line 1-7, but have now revised the text in both sections to make sure that this is clear. We have revised the end of section 4.2.1 to:

Signatures describing the frequency and duration of high and low flow events (Q_{HF}, Q_{HD}, Q_{LF}, and Q_{LD}) had large uncertainties in both catchments (±10–35%). This arises because the event threshold is defined as a multiplier of the mean or median flow, and so the (uncertain) gradient of the rating curve greatly impacts on the flow percentile equivalent to the threshold value. Frequency and duration signatures have alternatively had the event threshold defined directly as a flow percentile (Kennard et al., 2010; Olden and Poff, 2003); we suggest this is preferable as those signatures were insensitive to the uncertainties analysed here, apart from sometimes small effects when using daily averages.

The beginning of the last paragraph of section 5.1 in the discussion was revised to:

Signatures can be designed to be robust to some data uncertainty sources. A clear example is for signatures describing the frequency and duration of high and low flow events. If these events are defined using a threshold defined as a multiplier of the mean or median flow, they are highly sensitive to rating curve uncertainty. If instead, the events are directly defined using a flow percentile threshold, they were little affected by rating curve uncertainty (see Section 4.2.1).

Figs. 1 and 2: I think that the reader would get more understanding on the two study areas if the Authors would add a sample of the hydrograph in the figures (or in an additional one). This would show how the runoff responses differ in the two catchments (e.g., difference in flashiness). I am thinking to something like Figure 1 in http://www.hydrol-earth-syst-sci.net/17/2263/2013/hess-17-2263-2013.pdf

Best regards, Alberto Viglione

Response: Thank you for this useful suggestion, we have added a figure (Fig. 7 in the revised manuscript, see below) describing the two hydrographs and the runoff response, as well as more text describing the differences in runoff response in the end of Section 4.1.2, see below:

Mahurangi has a fast rainfall-runoff response with little base flow and peak flow events that are infrequent but have large magnitudes (up to 11 mm/h, Fig. 7a, right inset plot). Brue, by contrast, has higher base flow and more peak flow events of longer duration and lower magnitudes (up to 1 *mm/h, Fig 7b, right inset plot). Large high-flow uncertainty is likely in catchments such as the Mahurangi where peak flows occur seldom and last only a few hours – this makes reliable high-flow gauging practically difficult and rating-curve extrapolation often necessary. The larger high-flow rating-curve uncertainty in Mahurangi (Fig. 6a) is reflected in a wider peak flow uncertainty distribution (Fig. 7a, left inset plot). In Brue the whole flow range is gauged and the high-flow rating-curve uncertainty is smaller (Fig. 6c), the peak flow distribution has higher kurtosis with heavier tails (Fig. 7b, left inset plot).*

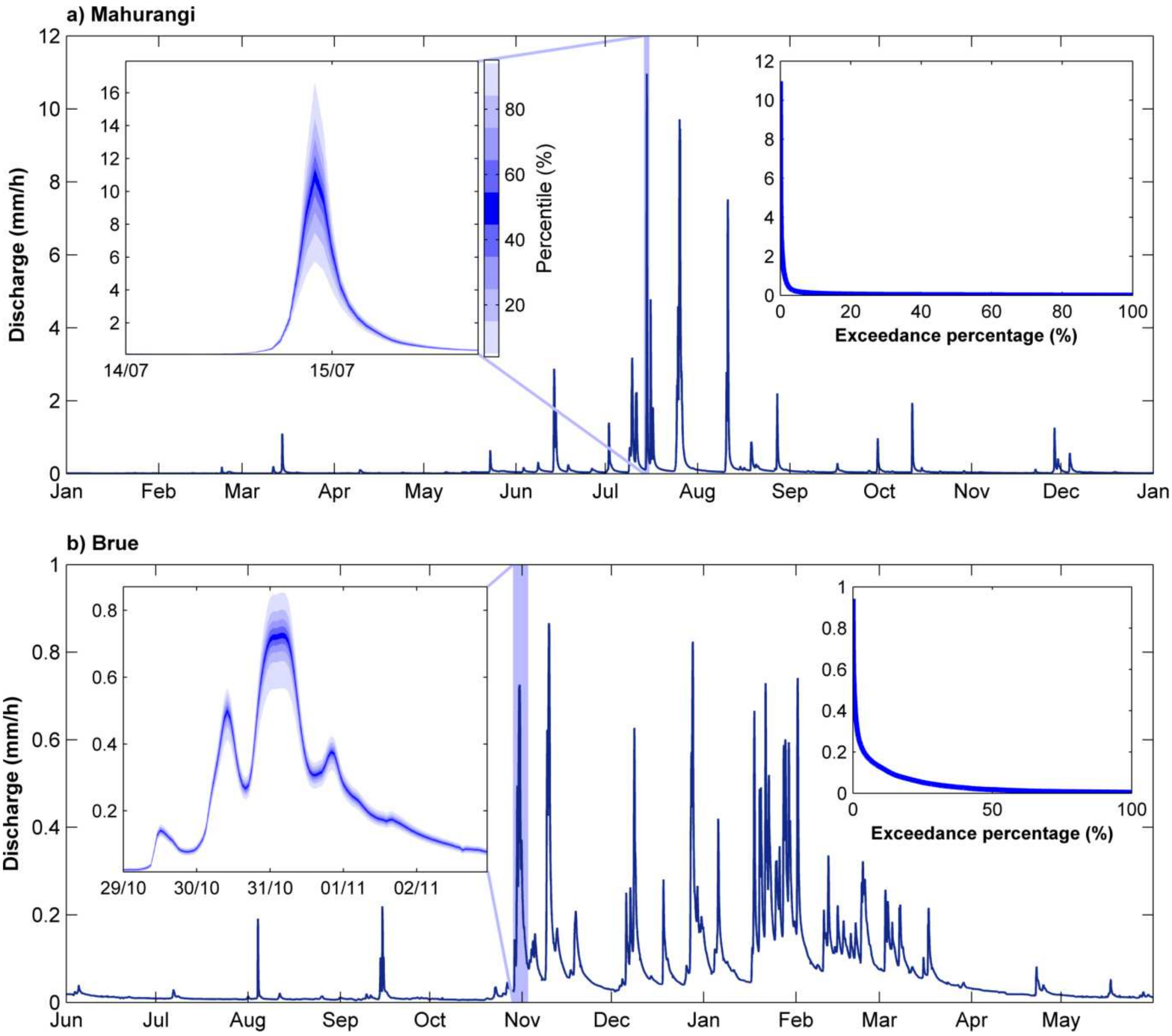

Fig. 7 Discharge calculated using the optimal rating curve for 1998 for Mahurangi (a) and for 1994–1995 for Brue (b). The left inset plots show the discharge time series uncertainty distribution at an hourly scale for a peak flow event in each catchment. The right inset plots show the flow-duration curves for the full time series for each catchment. The y-axis variable and unit is discharge in mm/h in all plots.

Response to Referee 2, H. Gupta

Referee Comments (Hoshin Gupta) on “Uncertainty in hydrological signatures by IK Westerberg and HK McMillan" submitted to HESS

I. Contributions of the Paper

A) Goals: (1) To contribute to awareness of signature uncertainty, including typical sources, magnitudes and methods for assessment. (2) To propose a general method for estimating signature uncertainty. (3) To demonstrate how typical uncertainty estimates translate to magnitude and distribution of signature uncertainty in two example catchments.

B) Summary: A diagnostic hydrological signature quantifies information from observed data as an index value. Uncertainties in the observed data, and subjective choices in the calculation method, propagate into the signature values and reduce their information content. However, uncertainty sources and distributions are application-specific, making a general analytic solution for signature uncertainty difficult. This paper reviews the uncertainties relevant to different signatures in rainfall and flow data, and proposes that a Monte Carlo simulation can provide a generally applicable and flexible method, by sampling equally likely possible realizations of the true data values, conditioned on the observed data (where multiple data sources are needed, grouped samples are used). Each realization is then used to calculate the signature value, and the values collated to give the signature distribution. Results are demonstrated for two catchments.

C) Findings: 1) Uncertainties are often large (±10–40% relative uncertainty) and highly variable between signatures. 2) Greater uncertainty in signatures that use highfrequency responses, small data subsets, or subsets prone to measurement errors. 3) Lower uncertainty in signatures that use spatial or temporal averages. 4) Some signatures are sensitive to particular uncertainty types such as rating-curve form.

D) Conclusions: Signatures can be designed to be robust to some uncertainty sources. Signature uncertainties of the magnitudes found have the potential to change the conclusions of hydrological and ecohydrological analyses, such as cross-catchment comparisons or inferences about dominant processes.

II. Referee Comments (Hoshin Gupta): This is a very well conceived and written paper. The organization and presentation are excellent. The subject matter is both timely and addressed in a clear and comprehensive manner. I recommend publication with no reservations.

Since I am not very well versed in the sources and nuances of observation/data uncertainty, I focused my review my attention mainly on the methodology applied. In general I concur that the Monte-Carlo approach is a suitable way to approach the problem of estimating signature uncertainty (and is more generally applicable in the context of data assimilation – i.e., estimating attributes of a dynamical systems model from data). The key sensitivity of the results will, of course, be to the choice of sampling distribution, and a certain amount of subjectivity is necessarily involved therein.

I commend the authors on another noteworthy paper (in their growing list of excellent contributions to the literature). I wonder only if they might choose to comment on (perhaps in the conclusions) in more detail on how the inevitable subjectivity involved in choice/construction of the sampling distribution might influence any interpretations, and whether (perhaps) the use of maximal entropy forms of sampling distributions (conditional, of course, on the actual data and what is qualitatively known), might help in this regard.

Response: We thank Hoshin Gupta for his very positive and kind comments about our paper. There is certainly some subjectivity in the choice of the uncertainty estimation methods and these should be motivated by the perceptual understanding of the uncertainty sources. We agree that this is an important consideration and are currently involved in a comparison study that aims to compare and better understand the effects of assumptions and methodological choices when it comes to estimating discharge uncertainty. We mentioned this issue briefly at the end of the introduction (P4237, line 13-23) to draw attention to this at the start of the paper. However, we agree with all the reviewers that it would be good to have some further discussion at the end of the paper and have therefore introduce an extra paragraph in the discussion in section 5.2 (see below).

With regards to the choice of sampling distribution, an example is the specification of the prior parameter distributions for the rating-curve parameters that can influence the results in the estimation of rating-curve uncertainty with the MCMC Voting Point method. This occurs primarily when the rating curve is extrapolated to ungauged flow levels, in particular if there are few gaugings in the high-flow section of the rating curve so that the prior distribution plays a larger role. In effect the prior distribution is playing the role of a perceptual model that brings information to the estimation problem, as Hoshin Gupta has discussed in recent papers – we have noted this in the discussion. This addresses an epistemic type of uncertainty related to lack of knowledge about the true stage-discharge relation that might be reduced by introducing new information (e.g. about the river cross-section and its characteristics) to constrain the uncertainty magnitudes. We have included some further discussion about this in the revised discussion (see new discussion paragraph below)

We recognise that the inferred distributions of signature uncertainty will be sensitive to the assumptions and methods used to estimate distributions of data uncertainty. This introduces some subjectivity into the uncertainty estimation and it is therefore important to make the assumptions explicit and motivate method choices by the perceptual understanding of the uncertainty sources. For example, the optimal methods for estimating rating curve uncertainty under typical time-varying, poorly-specified errors remain an active debate in the hydrological community. The use of an informal likelihood, as we did, rather than a formal statistical likelihood can be more robust to multiple epistemic error sources, but can also be criticised for not obeying a formal statistical framework (as discussed by McMillan and Westerberg (2015) and Smith et al. (2008)). Future progress in understanding how perceptual models and data jointly contribute to system identification may help to resolve this dichotomy (Gupta and Nearing, 2014). At present, we recognise that uncertainty distributions are more subjective in signatures that emphasise poorly-described aspects of data uncertainty such as out-of-bank flows.

Response to Referee 3, N. Le Vine

The paper considers uncertainty in hydrological signatures due to errors/uncertainties in rainfall and discharge time series: 1) point measurement, spatial interpolation and equipment malfunctioning errors for rainfall (no systematic errors), and 2) uncertainty in stage-discharge relation (no stage time series uncertainty). MCMC sampling is employed to estimate signature uncertainties based on the time series uncertainty. The findings illustrate individual and combined contributions of the above rainfall and discharge uncertainties to the extent of signature uncertainty; and show that each uncertainty source, except for the rainfall point measurement uncertainty, contributes to a sizable signature uncertainty (for the selected signatures).

The paper is well thought-through and addresses an existing gap in uncertainty assessment for hydrological time series and its propagation into hydrological signatures. One important aspect that, in my opinion, the authors need to acknowledge and discuss is that the 'posterior' distribution of the rating curves is not strictly a statistical distribution, since the Voting Point likelihood it is based on is not a formal statistical likelihood. This has implications on the use of the MCMC sampling method as well as on the interpretation of the corresponding signature values as draws from probabilistic distributions.

Response: We thank Nataliya Le Vine for the review and the positive comments about our paper. We have clarified the section in the end of the introduction, before the paragraph presenting the objectives of the paper (P4237, line 20-23 in the original manuscript)

This paper was focused on signature uncertainty rather than data uncertainty; we stress that alternative data uncertainty assessment methods could be used where the perceptual understanding of the uncertainty sources is different.

We have also included a section in the discussion about the choice of uncertainty estimation methods, including a comment on the use of formal/informal likelihoods (see new text in the response to Referee 2, Hoshin Gupta above). Here, we have also referred to the further motivation for the choice and development of the Voting Point likelihood found in the technical note where the Voting Point method was first presented (McMillan and Westerberg, 2015, as referred in the paper).

Further, I would suggest specifying in the title which hydrological signature uncertainty is considered in the manuscript, as there are other uncertainty sources, e.g. due to the time period selection, due to regionalization in ungauged basins.

Response: We appreciate the suggestion, but prefer to keep the more simple title of the paper, as it would become rather complex to specify exactly which uncertainty sources are considered. See the new Table 1 described in the response to Referee 1, Alberto Viglione, for a summary of uncertainty sources to assist the reader.

Uncertainty in hydrological signatures

I. K. Westerberg[1,2] and H. K. McMillan[3]

[1]{Department of Civil Engineering, University of Bristol, Queen's Building, University Walk, Clifton, BS81TR, UK}

[2]{IVL Swedish Environmental Research Institute, P.O. Box 210 60, 10031, Stockholm, Sweden}

[3]{National Institute of Water and Atmospheric Research, P. O. Box 8602, Christchurch, New Zealand}

Correspondence to: I. K. Westerberg (ida.westerberg@bristol.ac.uk)

Abstract

Information about rainfall–runoff processes is essential for hydrological analyses, modelling and water-management applications. A hydrological, or diagnostic, signature quantifies such information from observed data as an index value. Signatures are widely used, including for catchment classification, model calibration and change detection. Uncertainties in the observed data – including measurement inaccuracy and representativeness as well as errors relating to data management – propagate to the signature values and reduce their information content. Subjective choices in the calculation method are a further source of uncertainty.

We review the uncertainties relevant to different signatures based on rainfall and flow data. We propose a generally applicable method to calculate these uncertainties based on Monte Carlo sampling and demonstrate it in two catchments for common signatures including rainfall-runoff thresholds, recession analysis and basic descriptive signatures of flow distribution and dynamics. Our intention is to contribute to awareness and knowledge of signature uncertainty, including typical sources, magnitude and methods for its assessment.

We found that the uncertainties were often large (i.e. typical intervals of ±10–40% relative uncertainty) and highly variable between signatures. There was greater uncertainty in signatures that use high-frequency responses, small data subsets, or subsets prone to measurement errors. There was lower uncertainty in signatures that use spatial or temporal averages. Some signatures were sensitive to particular uncertainty types such as rating-curve

form. We found that signatures can be designed to be robust to some uncertainty sources. Signature uncertainties of the magnitudes we found have the potential to change the conclusions of hydrological and ecohydrological analyses, such as cross-catchment comparisons or inferences about dominant processes.

1 Introduction

1.1 Hydrological signatures and observational uncertainty

Information about rainfall–runoff processes in a catchment is essential for hydrological analyses, modelling and water-management applications. Such information derived as an index value from observed data series (rainfall, flow and/or other variables) is known as a hydrological or diagnostic signature, and these are widely used in both hydrology (Hrachowitz et al., 2013) and ecohydrology (Olden and Poff, 2003). The reliability of signature values depends on uncertainties in the data and calculation method, and some signatures may be particularly susceptible to uncertainty. Signature uncertainties have so far received little attention in the literature; therefore guidance on how to assess uncertainty, and typical uncertainty magnitudes would be valuable.

Signatures are used to identify dominant processes and to determine the strength, speed and spatiotemporal variability of the rainfall-runoff response. Common signatures describe the flow regime (e.g. Flow Duration Curve, FDC, and recession characteristics), and the water balance (e.g. runoff ratio and catchment elasticity, Harman et al., 2011). Field studies have identified drivers of catchment function, such as a threshold response to antecedent wetness (Graham et al., 2010b; Penna et al., 2011; Tromp-van Meerveld and McDonnell, 2006a), which have been captured as signatures (McMillan et al., 2014). Signatures often incorporate multiple data types, including soft data (Seibert and McDonnell, 2002; Winsemius et al., 2009).

There is a long history of using flow signatures in eco-hydrology to assess instream habitat including the seasonal streamflow pattern, and the timing, frequency and duration of extreme flows (e.g. Jowett and Duncan, 1990). Signatures are used to detect hydrological change, e.g. Archer and Newson (2002) used flow signatures to assess the impacts of upland afforestation and drainage. Signatures can define hydrological similarity between catchments (McDonnell and Woods, 2004; Sawicz et al., 2011; Wagener et al., 2007), and assist prediction in ungauged basins (Bloeschl et al., 2013). Model calibration criteria using signatures are useful because they preserve information in measured data (Gupta et al., 2008; Refsgaard and Knudsen, 1996; Sugawara, 1979). Signatures used in calibration include the FDC (Westerberg et al., 2011), flow entropy (Pechlivanidis et al., 2012), the spectral density function (Montanari and Toth, 2007), or combinations of multiple signatures (Pokhrel et al., 2012). By using signatures that target individual modelling decisions, model components can be tested for compatibility with observed data (Clark et al., 2011; Coxon et al., 2013; Hrachowitz et al., 2014; Kavetski and Fenicia, 2011; Li and Sivapalan, 2011; McMillan et al., 2011). Hydrological signatures have been regionalised to ungauged basins and then used to constrain a model for the ungauged basin (Kapangaziwiri et al., 2012; Westerberg et al., 2014; Yadav et al., 2007).

Some previous authors have considered the effect of data uncertainty on hydrological signatures (Kauffeldt et al., 2013), particularly in model calibration. Blazkova and Beven (2009) incorporate uncertainties in signatures used as limits of acceptability to constrain hydrological models. Juston et al. (2014) investigate the impact of rating-curve uncertainty on FDCs and change detection for a Kenyan basin. They show that uncertainty in extrapolated high flows creates significant uncertainty in the FDC and the total annual flow. Kennard et al. (2010) discuss the uncertainties affecting ecohydrological flow signatures from measurement error, data retrieval and preprocessing, data quality, and the hydrologic metric estimation.

1.2 Uncertainty considerations relevant for hydrological signatures

We present a short description of data uncertainties relevant to hydrological signatures (see McMillan et al. (2012), for a longer review). In general, data uncertainties stem from 1) measurement uncertainty (e.g. instrument inaccuracy or malfunction), 2) measurement representativeness for the variable under study (e.g. point rainfall compared to catchment average rainfall), and 3) data management uncertainty (e.g. data entry errors, filling of missing values or station coordinate errors). Errors from data management, equipment malfunction or human errors can often be detected and corrected in quality control (Bengtsson and Milloti, 2010; Eischeid et al., 1995; Viney and Bates, 2004; Westerberg et al., 2010). But some data errors, e.g. poorly calibrated or off-level raingauges, are difficult to correct post hoc (Sieck et al., 2007). The calculation of some signatures requires subjective decisions that introduce extra uncertainty, for example storm identification criteria, data time step, and whether to split the data by month/season (e.g. Stoelzle et al., 2013).

Each uncertainty component requires an error model that specifies the error distribution and dependencies (e.g. errors may be heteroscedastic and/or autocorrelated). It is essential that the error model accurately reflects the uncertainty, rather than simply adding random noise, as hydrological uncertainties are typically highly structured. Some measurement uncertainties can be estimated by repeated sampling, whereas representativeness errors are difficult to estimate. The latter are often epistemic due to lack of knowledge at unmeasured locations/time periods (e.g. rainfall distant from rain gauges). The most appropriate method to assess data uncertainty depends on the information available and the hydrologist's knowledge of the catchment. For example, the choice of likelihood function may depend on characteristics of the data errors and the measurement site. Uncertainty estimation depends on the perceptual understanding of the uncertainty sources as well as the studied system and there is potential for a false sense of certainty about uncertainty where strong error model assumptions are made (Brown, 2004). Juston et al. (2014) refer to *uncertainty²* and show how interpretation of uncertainties as random vs systematic affects hydrologic change detection. This~~e~~ ~~main aim of this~~ paper was ~~to study~~focused on signature uncertainty rather than data uncertainty; we stress that alternative data uncertainty assessment methods could be used where the perceptual understanding of the uncertainty sources is different.

The objectives of this paper were: (1) to contribute to the community's awareness and knowledge of observational uncertainty in hydrologic signatures, (2) to propose a general method for estimating signature uncertainty, and (3) to demonstrate how typical uncertainty estimates translate to magnitude and distribution of signature uncertainty in two example catchments.

2 Catchments and data

We used two catchments: the Brue catchment in the UK, and the Mahurangi catchment in New Zealand. This enabled us to compare signature uncertainties in different locations and with different uncertainty sources. Both catchments have excellent raingauge networks that allowed us to quantify uncertainty in rainfall data, and there is some existing knowledge of the dominant hydrological processes.

2.1 The Mahurangi catchment

The Mahurangi is a 50 km² catchment in the North Island of New Zealand. It has a warm and humid climate, with mean annual rainfall of 1600 mm yr^{-1}. The catchment has hills and gently rolling lowlands, and land use is a mixture of pasture, native forest and pine plantation. The soils are clay loams, less than 1 m deep. Extensive datasets of rainfall and flow were collected during the Mahurangi River Variability Experiment 1997–2001 (Woods et al., 2001). We used hourly data from the 13 tipping bucket rain gauges and the catchment outlet flow gauge for 1 January 1998–31 December 2000 (Fig.1). Missing rainfall values were available from a previous study that had infilled them using linear correlation with a nearby site. The flow gauge has a two-part triangular weir for low to medium flows, and a rated section with confining wooded banks for high flows. During the study period, the maximum recorded stage was 3.8 m, but the highest gauged stage is 2.7 m.

2.2 The Brue catchment

The predominantly rural 135 km^2 Brue catchment in south-west England has low grassland hills of up to 300 m above sea level (Fig. 2). Clay soils overlay alternating bands of permeable and impermeable rocks. An extensive precipitation dataset consisting of 49 tipping-bucket raingauges and radar data with 15-min resolution was created by the HYREX (Hydrological Radar Experiment) project (Moore, 2000; Wood et al., 2000). We used the data from 1 January 1994 to 31 December 1997, with a mean annual precipitation of 820 mm yr^{-1}. The extensive quality control described by Wood et al. (2000) included analyses of monthly cumulative rainfall totals and correlation analyses of timing errors. Errors included instrument malfunctions such as funnels blocked by debris and mouse damage to electrical cables. There were thus substantial periods of missing data resulting after quality control (Fig. 2), even for these carefully maintained rain gauges. We interpolated the missing precipitation values with inverse-distance weighting to obtain a complete dataset for subsampling analysis.

The Lovington discharge station has a crump profile weir for low flows and a rated section above 2.2 m^3/s. The whole stage range was gauged and the water was below bankfull level for the chosen period. The stage-discharge relationship is affected by downstream summer weed growth resulting in scatter in the low-flow part of the rating curve.

3 Method: Estimation of uncertainty in hydrological signatures

Uncertainty sources and distributions are application-specific, so a general analytic solution for the signature uncertainty is not available. We suggest that Monte Carlo simulation provides a generally applicable and flexible method, by sampling equally likely possible realisations of the true data values (e.g. rainfall or flow series), conditioned on the observed data. Where multiple data sources are needed (e.g. calculation of runoff ratio), paired samples are used. Each sampled data series is used to calculate the signature value, and the values collated to give the signature distribution. This technique has previously been used to determine uncertainty in discharge (McMillan et al., 2010; Pappenberger et al., 2006) and rainfall (Villarini and Krajewski, 2008).

We applied the Monte Carlo (MC) approach to estimate uncertainty in signatures of different complexity. We used signatures that require rainfall and/or streamflow data only. Our method is described in Fig. 3 and has four steps: 1) identification of uncertainty sources in the data and from subjective decisions in signature calculation, 2) specification of uncertainty models for each uncertainty source either from the literature or catchment-specific analyses, 3) Monte Carlo sampling from the different uncertainty models and calculation of signature values for each sample, and 4) analyses of the estimated signature distributions, their dependence on individual uncertainty sources and comparisons between catchments. We analysed both the absolute and relative uncertainty distributions, where the relative uncertainties were defined using the signature value from the best-estimate discharge and precipitation.

3.1 Method: Data uncertainty sources and their estimation

We first describe the error models for uncertainties relating to rainfall and flow. Further uncertainty sources that are specific to a particular signature are described separately in Section 3.2. Table 1 presents a summary of all uncertainty sources together with literature references for the uncertainty estimation methods.

3.1.1 Catchment average rainfall

Identification of uncertainty sources

We considered catchment average rainfall estimated from a network of rain gauges, with three main uncertainty sources: point measurement uncertainty, spatial interpolation uncertainty and equipment malfunction uncertainty (e.g. unrecognised blocked gauges). Point uncertainty includes random errors such as turbulent airflow around the gauge (Ciach, 2003), and is usually assessed using co-located gauges. Systematic point errors are also common (e.g. undercatch due to wind loss, wetting loss, splash-in/out). In theory, systematic errors can be corrected for, but this is difficult and the site-specific information required is not always available (Sieck et al., 2007). In this study, we considered random point uncertainty but not systematic components. Interpolation errors occur when estimating catchment average rainfall from the point measurements at the gauges and depend on rainfall spatial variability (affected by topography, rain rate and storm type), density of gauges and network design.

Uncertainty estimation method

Point uncertainty was calculated using the formula derived by Ciach (2003) from a study of 15 co-located tipping bucket rain gauges over 12 weeks:

$$\sigma = 0.0035 + 0.2/r \qquad (1)$$

Where r is the rainfall rate in mm/hr and σ is the standard deviation of the relative error in 1 hour measurements. No information about the distribution of the errors was given; we assumed a Gaussian distribution with zero mean. Interpolation uncertainty was estimated by sub-sampling from the gauge network. We subsampled using 1–13 (1–49) gauges for Mahurangi (Brue) for the basic signatures. For the combined rainfall-runoff signatures, 3 gauge densities were used: 1 gauge/45km^2, 1 gauge/10km^2 and 1 gauge/5km^2, which equalled 1 (3), 5 (14) and 10 (28) gauges in Mahurangi (Brue) respectively. We also used the single gauge case for Brue. Each subsampled dataset was used to estimate areal average rainfall at each time step using Theissen polygon interpolation. Equipment malfunction uncertainty was investigated for Brue, where a quality-assured set of reliable periods was available (Section 2.2). We repeated our analyses using both the raw and quality-controlled data sets.

3.1.2 Discharge data

Identification of uncertainty sources

We considered discharge as estimated from a measured stage series and a rating curve that relates stage to discharge. This is the most common method, and is used at both our case study sites. The main uncertainty sources are:

(1) Uncertainty in the gaugings (i.e. the measurements of stage and discharge used to fit the rating curve). Discharge uncertainty is typically larger, but during high flow gaugings, stage can change rapidly and its average may be difficult to estimate.

(2) Approximation of the true stage-discharge relation by the rating curve. This is usually the dominant uncertainty (McMillan et al., 2012), especially when the stage-discharge relation changes over time. In both catchments, low to medium flows are contained within a weir, which constrains the uncertainty. However, for Brue considerable low-flow uncertainty remains as a consequence of seasonal vegetation growth.

Uncertainty in the stage time series was not assessed apart from correcting obvious outliers. For Brue, occasional periods where stage data had been interpolated linearly from lower-frequency measurements were excluded from the recession analysis.

Uncertainty estimation method

We used the Voting Point likelihood method to estimate discharge uncertainty by sampling multiple feasible rating curves (McMillan and Westerberg, 2015). In brief, discharge gauging uncertainty was approximated by logistic distribution functions based on an analysis of 26 UK flow gauging stations with stable rating sections (Coxon et al., ~~In review~~2015). This analysis gave 95% relative error bounds of 13–14 % for high flow to 30–40% for low flow (noting that the logistic distribution is heavy-tailed). Stage gauging uncertainty was approximated by a uniform distribution of ±5 mm, a mid-range value based on previous studies (McMillan et al., 2012).

Rating-curve uncertainties, including extrapolation and temporal variability, were jointly estimated using Markov Chain Monte Carlo (MCMC) sampling of the posterior distribution of rating curves consistent with the uncertain gaugings. The Voting Point likelihood draws on previous methods that account for multiple sources of discharge uncertainty (Juston et al., 2014; Krueger et al., 2010; McMillan et al., 2010; Pappenberger et al., 2006). The rating curve forms were based on the official curves, where Mahurangi had a 3-segment power law curve and Brue a 2-segment (for the range of flows analysed here). The power law parameters and the breakpoints were treated as parameters for estimation.

3.2 Method: Calculation of hydrological signatures with uncertainty

3.2.1 Basic signatures

A set of signatures describing different aspects of the rainfall-runoff behaviour were calculated (Table 2~~1~~). We used signatures describing flow distribution, event characteristics, flow dynamics and rainfall; flow timing would be less affected by the data uncertainties studied here. Only data uncertainty (i.e. no subjective decisions) was considered for the basic signatures.

3.2.2 Recession analysis

Recession analysis is widely used to study the storage-discharge relationship of a catchment (Hall, 1968; Tallaksen, 1995), which gives insights into the size, heterogeneity and release characteristics of catchment water stores (Clark et al., 2011; Staudinger et al., 2011). We used the established method of characterising the relationship between flow and its time-derivative. In the theoretical case where flow Q is a power function of storage, and evaporation is negligible, the relationship is:

$$d\hat{Q}\big/dt = -\hat{Q}^{b}\big/T_0 \qquad (2)$$

Where $\hat{Q} = Q/Q_0$ is flow scaled by the median flow Q_0. T_0 and b are found by plotting $-dQ/dt$ against Q on logarithmic axes. T_0 is the characteristic recession time at the median flow. b indicates nonlinearity of response: $b = 1$ implies a linear reservoir, $b > 1$ implies greater nonlinearity or multiple water stores with different drainage rates (Clark et al., 2009; Harman et al., 2009).

Subjective decisions in recession analysis include how recession periods are defined, the delay after rainfall used to eliminate quickflow, the data time step, and whether to extend time steps during low flows to improve flow derivative accuracy (Rupp and Selker, 2006). A moving average can be used to smooth diurnal flow fluctuations. Options to estimate T_0 and b include linear regression, total least squares regression to allow for errors in both variables (Brutsaert and Lopez, 1998), or regression on binned data values (Kirchner, 2009). If water distributions vary seasonally, the results are sensitive to whether recessions are fitted using all data combined – or split by season, month or event (Shaw and Riha, 2012).

We assessed subjective uncertainty in recession analysis by comparing the distributions of recession parameters *b* and T_0 in the following cases, which in our experience have the most potential to affect recession parameter values: (1) using hourly vs daily flow data, and (2) calculating recession parameters using all data combined vs calculating parameters by season and taking the mean.

3.2.3 Thresholds in rainfall-runoff response

Threshold behaviour in the relationship between rainfall depth and flow contributes to hydrological complexity (Ali et al., 2013) and exerts a strong control on model predictions. Threshold identification depends on both rainfall and flow data, making it a good candidate to test the effect of multiple uncertainty sources. Rainfall-runoff thresholds have been found in many catchments (Graham et al., 2010b; Tromp-van Meerveld and McDonnell, 2006a, b) including the Mahurangi (McMillan et al., 2011; McMillan et al., 2014). We only studied threshold signatures in the Mahurangi, as the Brue did not display any rainfall-runoff threshold.

The signatures that we used were threshold location (in mm of rain per event) and threshold strength. We quantified threshold strength based on the method of McMillan et al. (2014). Storm events were identified and event rainfall was plotted against event runoff. Strong threshold behaviour was defined as an abrupt increase in slope of the event rainfall-runoff relationship. This attribute was tested by fitting each data set with two intersecting lines (a 'broken stick' fit), using total least squares to optimise the slopes and intersect. The corresponding null hypothesi~~e~~s was that the two lines have equal slopes. This test returns a z-statistic which quantifies the strength of evidence for the alternative hypothesis: where the absolute value exceeds 1.96, the null hypothesis can be rejected at the 5% level.

We defined events based on McMillan et al. (2011), such that events require at least 2 mm/hour or 10 mm/day of precipitation, and are deemed to end either when a new event begins, or five days after the last rainfall. Events are distinct if they are separated by 12 dry hours. We assessed uncertainty due to subjective decisions by using or not using baseflow separation, and by changing the event definition to include smaller events, where at least 1 mm/hour or 5 mm/day of precipitation fell. We used the baseflow separation method of Gustard et al. (1992), which interpolates linearly between 5-day flow minima to create the baseflow series.

4 Results

4.1 Estimated uncertainty in rainfall and discharge data

4.1.1 Rainfall data

The standard deviation of the error in catchment average rainfall resulting from different numbers of subsampled stations was calculated. It was plotted as a function of hourly rain rate using the moving-average window method of Villarini and Krajewski (2008), with a bandwidth equal to 0.7 times the rain rate at the centre of the window (results for the Brue in Fig. 4). The errors decreased with rain rate and there was a large initial decrease in the error when the number of sub-sampled stations increased from 1 to around 5. The point uncertainty only had a small effect on the error standard deviation.

The number of gauges had a large effect on the estimated mean annual precipitation; if only one rain gauge was used, there was a range of 200–300 mm yr^{-1} that would clearly affect catchment water balance analyses (Fig. 5). One rain gauge in a catchment of this size is still well above the WMO recommended station density of 1 gauge per 575 km^2 in hilly terrain (WMO, 2008). Here there was also a large initial decrease in the range when the number of gauges increased to around five. But, even when three or four gauges were used (1 gauge per 12–16 km^2) for Mahurangi, there was a 1430–1660 mm yr^{-1} ~~uncertainty~~ range in mean annual precipitation. When the non-quality controlled dataset was used for the Brue (Fig. 5a), there was a decrease in both mean annual values and standard deviation. At the same time the range in standard deviation increased because stations with erroneously high or missing precipitation values were retained (blocked rain gauges were a particular problem in this catchment; Wood et al., 2000). The estimated precipitation standard deviation was uncertain for one subsampled gauge in the Mahurangi (Fig. 5c), where gauges were located in both the wettest and driest parts of the catchment.

4.1.2 Discharge data

The estimated rating-curve uncertainty is shown in Fig. 6, with the corresponding flow percentile uncertainty summarised using boxplots. The 5–95 percentile uncertainty bounds enclose almost all of the uncertain gaugings, apart from a small number of outliers. Low flow uncertainty is larger in Brue where vegetation growth affects the stability of the stage-discharge relation. High flow uncertainty is larger in Mahurangi where fewer, more scattered high flow gaugings cause a wider range in the extrapolated flows. Mahurangi has a fast rainfall-runoff response with little base flow and peak flow events that are infrequent but have large magnitudes (up to 11 mm/h, Fig. 7a, right inset plot). Brue, by contrast, has higher base flow and more peak flow events of longer duration and lower magnitudes (up to 1 mm/h, Fig 7b, right inset plot). Large high-flow uncertainty is likely in catchments such as the Mahurangi where peak flows occur seldom and last only a few hours – this makes reliable high-flow gauging practically difficult and rating-curve extrapolation likely necessary. The larger high-flow rating-curve uncertainty in Mahurangi (Fig. 6a) is reflected in a wider peak flow uncertainty distribution (Fig. 7a, left inset plot). In Brue the whole flow range is gauged and the high-flow rating-curve uncertainty is smaller (Fig. 6c), the peak flow distribution has higher kurtosis with heavier tails (Fig. 7b, left inset plot).

4.2 Estimated uncertainty in the hydrological signatures

4.2.1 Basic signatures

Flow percentile uncertainties mirrored those of the rating curves, with larger uncertainties in high-flow percentiles for Mahurangi and larger uncertainties in low-flow percentiles for Brue (Fig. 6). Uncertainty in mean discharge was around ±10% for both catchments; this is the 5–95 percentile interval, the distributions are shown in Fig. 8~~7~~. Signatures describing the flow variability (S_{FDC}, Q_{CV}, and Q_{AC}) had much higher uncertainties in Mahurangi (±20–50%), where there was a fast rainfall–runoff response and greater high-flow rating uncertainty. The uncertainty in the S_{FDC} was particularly large for Mahurangi because the rating curve had a breakpoint in the 33–66 percentile interval used to calculate the slope. Signatures describing the frequency and duration of high and low flow events (Q_{HF}, Q_{HD}, Q_{LF}, and Q_{LD}) had large uncertainties in both catchments (±10–35%). This arises because the event ~~defined with a~~ threshold is defined as a multiplier of the mean ~~and~~ or median flow, and so the (uncertain) gradient of the rating curve greatly impacts on the flow percentile equivalent to the threshold value~~ had large uncertainties in both catchments (±10–35%)~~. Frequency and duration signatures have alternatively ~~been defined using~~ had the event threshold defined directly as a flow percentile (Kennard et al., 2010; Olden and Poff, 2003); we suggest this is preferable as those signatures were insensitive to the uncertainties analysed here, apart from sometimes small effects when using daily averages.

4.2.2 Total Runoff Ratio

For the total runoff ratio, we tested the contribution of each uncertainty source by including or excluding different sources. We calculated total uncertainty (Fig 8~~7~~c-d, black bars) using different rain-gauge densities. Total uncertainty was approximately ±15% using a single rain gauge, decreasing slowly with more gauges. The distributions were largely unbiased when using quality-controlled data. The contribution of point precipitation uncertainty was minimal: excluding this source made no difference to the uncertainty distribution (Fig 8~~7~~, green bars). Precipitation uncertainty is therefore due to interpolation, and was evaluated by excluding flow uncertainty and calculating the remaining uncertainty (Fig 8~~7~~, blue bars). This uncertainty was noticeable (approximately ±10% Mahurangi, ±9% Brue) for one gauge but decreased quickly with more gauges and was negligible at a density of 1 gauge per 5 km^2. Total uncertainty was dominated by discharge uncertainty (dark blue bars) which was greater than precipitation uncertainty (blue bars). In the Brue catchment the effect of using un-quality controlled data was assessed (red and purple bars) which increased and biased the uncertainty, particularly at low gauging densities.

4.2.3 Recession analysis

We tested the effect of data uncertainty on recession analysis results by plotting histograms of the recession parameters b (nonlinearity of recession shape) and T_0 (recession slope at median flow). We considered subjective uncertainty by using data at daily or hourly time steps, and calculating parameters using all data together or splitting by season and then taking parameter averages (Fig. 9~~8~~).

Uncertainty in the recession descriptors was typically (1) greater for Brue than for Mahurangi, in particular for hourly flow data; (2) greater for hourly flow data than for daily flow data. Recessions are calculated from flow derivatives, and are therefore affected by relative changes in flow (e.g. channel shape). The linear regression used to calculate the recession parameters is particularly sensitive to uncertainties in extreme low or high flows. The low flow uncertainty at Brue resulting from summer weed growth creates higher uncertainties at that site. Daily flow values are based on an aggregation of measured values, and are therefore more robust to data uncertainty. However, using daily data in small catchments can mask details of the recession shape, as the slope can change markedly during a single day. In our case, this difference caused shifts in the parameter distributions between hourly and daily data, and would therefore affect our ability to compare parameter values between catchments. For example, b values were similar in the two catchments when using daily data, but different when using hourly data; and the converse is true for T_0. This was caused by differences in the hydrograph such as low flow fluctuations in the Brue and flashy peak-flow events in the Mahurangi.

Recession parameters calculated per season were highly uncertain in the Brue for the T_0 parameter. This was due to some seasons having very few recession data points, and therefore the fitted regression relationships being sensitive to changes in these points. Recession parameters were highly sensitive to subjective decisions in defining recession periods, as also found by Stoelzle et al. (2013). Such definitions could result in particular recession periods being included or excluded from the analysis depending on the sampled rating curve. When the excluded periods included extreme high or low flow values, this could significantly skew the fitted parameters, and therefore give multimodal parameter distributions according to the particular set of valid recession periods. For the daily timescale, the starting hour used in calculating the daily averages could also have a large effect on the resulting recession parameters.

4.2.4 Thresholds in rainfall-runoff response

We tested for uncertainty in the estimated threshold in the event rainfall-runoff relationship in Mahurangi using box plots of the threshold location and strength under different uncertainty scenarios (Fig. 10~~9~~). The threshold broken-stick fit is illustrated in Fig. 10~~9~~a for the best-estimate data (in blue) and for an example realisation with uncertainty (in grey).

The threshold was 65 mm when using best-estimate rainfall and flow data. Total uncertainty was a largely unbiased distribution with a range of ~20 mm. Total uncertainty was a combination of flow uncertainty (slight low bias) with rainfall interpolation uncertainty (slight high bias). Point rainfall uncertainty was not important when using multiple gauges. Threshold location was highly sensitive to the number of rain gauges used: using only one gauge created a very wide uncertainty distribution. As with the rainfall uncertainty analysis, there was a large decrease in the uncertainty when increasing to five gauges (Section 4.1.1). The use of baseflow separation did not greatly change the median threshold, but did increase the range. Event definition parameters had little effect on the threshold uncertainty.

Threshold strength was defined using a change-in-slope statistic where higher values indicate a stronger threshold. Considering flow or rainfall uncertainty weakened the calculated threshold. For flow uncertainty this was due to the optimal rating curve having its first breakpoint and mid-section slope above the median values of the sampled rating curve distribution; both of which were associated with a stronger threshold. As with the S_{FDC} this shows the strong impact of the rating curve breakpoint locations on signature uncertainty. For rainfall, uncertainty adds noise to the event rainfall depth and therefore corrupts the estimated rainfall-runoff relationship, weakening the threshold. Consequently, the number of rain gauges is an important control on estimated threshold strength, with fewer gauges causing a weakened threshold. As the underlying threshold was strong, the case of 1 rain gauge was the only scenario that could cause the threshold statistic not to be significant at the 5% level. However, in other catchments with weaker thresholds, lack of good rainfall data is likely to result in thresholds being missed. Using baseflow separation increased the derived threshold strength, as it typically reduced runoff depths for smaller events below the threshold. Event definition had only a small effect on derived threshold strength; when smaller events were included the threshold strength statistic increased, as the fit was based on a greater number of points.

4.3 Summary of the signature uncertainties

To summarise our results, we tabulated examples of each signature type together with their dominant uncertainty sources and summary statistics of the total uncertainty distribution, for each catchment (Table 3). Our aim is to allow easy comparison of the signature uncertainties in our study, with those of other studies. We therefore chose commonly used distribution statistics, i.e. the first three distribution moments (mean, standard deviation, skewness); and the half-width of the 5–95 percentile range, which is commonly quoted in uncertainty studies (e.g. McMillan et al, 2012). We hope that authors of future studies will consider using similar statistics, to enable the community to compile a generalised understanding of signature uncertainties across different catchments, scales and landscapes.

5 Discussion

5.1 Uncertainty in different types of signatures

Uncertainty distributions were highly variable between signatures and therefore the impact of the uncertainty depends on which signatures are used (Table 3). There was greater uncertainty in signatures that use high-frequency responses (e.g. variations over short timescales, thresholds based on event precipitation totals), subsets of data more prone to measurement errors (e.g. extreme high and low flows, Q_{HV} and Q_{99}), and signatures based on small numbers of values (e.g. seasonal recession characteristics in the Brue). Signatures describing flow variability were uncertain in the Mahurangi catchment that has a flashy rainfall-runoff response and where stage significantly exceeded the highest gaugings leading to large discharge uncertainty at high flows. This is likely to be a common situation in small, fast-responding catchments with few high flow events, due to the practical difficulties of gauging during such short time-windows. There was lower uncertainty in signatures that use spatial or temporal averages (e.g. total runoff ratio and BFI). Uncertainty in signatures calculated from averages depends on the type of data uncertainty, e.g. random errors are reduced by averaging, but some systematic errors such as rainfall undercatch are not. Rating-curve uncertainty is an intermediate case as it depends on error magnitudes that vary across the flow range. Some signatures are sensitive to particular types of data uncertainty. For example in Mahurangi, high uncertainty in S_{FDC} relates to uncertainty in rating curve shape, and in Brue, high uncertainty in Q_{LD} relates to uncertainty of the low flow rating in combination with the shape of the hydrograph. Signatures that describe the rainfall–runoff relationship for individual events (e.g. threshold location and strength) were particularly sensitive to precipitation uncertainties for low gauging densities.

Signatures can be designed to be robust to some data uncertainty sources. A clear example is for signatures describing the frequency and duration of high and low flow events. If these events are defined using a threshold ~~related~~ defined as a multiplier of~~to~~ the mean or median flow, they are highly sensitive to rating curve uncertainty. If instead, the events are directly defined using a flow percentile threshold, they were little affected by rating curve uncertainty (see Section 4.2.1). This simple change in signature definition reduces sensitivity to data uncertainty. We found that any cut-offs imposed in signature calculation, such as event or recession definition criteria, could have a strong and unpredictable effect on signature uncertainty. For example, rainfall-runoff threshold strength calculations were particularly sensitive to large storm events, which control the gradient of the second line in the 'broken stick'. If such events were conditionally excluded (e.g. classified as disinformative and removed when runoff exceeded rainfall; which depends on the rating curve and raingauge(s) selected), the resulting uncertainty could overwhelm any other uncertainty sources. We suggest that signatures including cut-off type definitions should be carefully evaluated, and the cut-offs removed if possible.

5.2 Method limitations and future developments

The quality of signature uncertainty estimates relies on accurate assessment of data uncertainty and therefore in turn on sufficient information. An example of insufficient uncertainty information would be for a gauge where out-of-bank flows occur, but there is no information on the out-of-bank rating. As discussed by Juston et al. (2014) for rating curve uncertainty, it is essential to understand whether data errors are random or systematic; aleatory or epistemic. In our study, point rainfall errors were not important in signature uncertainty, but there is scope to improve their representation as systematic or random (e.g. systematic wind-related undercatch, or random turbulence effects). However, quantification of these errors is not straightforward (Sieck et al., 2007).

We recognise that the inferred distributions of signature uncertainty will be sensitive to the assumptions and methods used to estimate distributions of data uncertainty. This introduces some subjectivity into the uncertainty estimation and it is therefore important to make the assumptions explicit and motivate method choices by the perceptual understanding of the uncertainty sources. For example, the optimal methods for estimating rating curve uncertainty under typical time-varying, poorly-specified errors remain an active debate in the hydrological community. The use of an informal likelihood, as we did, rather than a formal statistical likelihood can be more robust to multiple epistemic error sources, but can also be criticised for not obeying a formal statistical framework (as discussed by McMillan and Westerberg (2015) and Smith et al. (2008)). Future progress in understanding how perceptual models and data jointly contribute to system identification may help to resolve this dichotomy (Gupta and Nearing, 2014). At present, we recognise that uncertainty distributions are more subjective in signatures that emphasise poorly-described aspects of data uncertainty such as out-of-bank flows.

For signatures calculated over a long time period, it may be appropriate to incorporate non-stationary error characteristics, such as rating curve shifts or the example explored by Hamilton and Moore (2012) where the best-practice method for infilling discharge values under ice changed over time. The time period used is important if signatures are used for catchment classification: an unusual event such as a large flood may shift the signature values (Casper et al., 2012). Additional uncertainty sources can be important in other catchments, such as catchment boundary uncertainty and flow bypassing the gauge (Graham et al., 2010a).

5.3 Implications for use of signatures in hydrological analyses

Our results are pertinent to any hydrological analysis that uses signatures to assess catchment behaviour. Examples of applications whose reliability could be affected by signature uncertainty include: testing bias correction of a climate model using signatures in a coupled hydrological model (Casper et al., 2012), predicting signatures in ungauged catchments (Zhang et al., 2014), classifying catchments using flow complexity signatures (Sivakumar et al., 2013), or assessing spatial variability of hydrological processes (McMillan et al., 2014). In some cases, absolute signatures values are not used, rather it is the pattern or gradient over the landscape, or trend over time that is important. Data uncertainties may obscure such patterns depending on the magnitude of the uncertainty in relation to the strength of the measured pattern. The range of signature values found by McMillan et al. (2014) across Mahurangi was large compared to the uncertainty magnitudes found in this study. This suggests that the conclusions regarding the signature patterns would still hold, assuming that the uncertainty at the catchment outlet is representative for the internal subcatchments. Some subjective uncertainty sources may not be relevant in catchment comparisons, as choices such as how to define recession periods or whether to do baseflow separation can be chosen consistently. However, subjective uncertainties can still change the conclusions drawn such as the cut-offs described above, and as discussed in section 4.2.3 where daily data suggested similar recession *b* parameters in Mahurangi and Brue, but hourly data showed strong differences.

When signatures are used as a performance measure in model calibration (e.g. Blazkova and Beven, 2009) reliable uncertainty estimates are crucial so that the model is not overfitted. Previous studies have quantified data and signature uncertainty using upper and lower bounds (e.g. fuzzy estimates used by Coxon et al., 2013; Hrachowitz et al., 2014; Westerberg et al., 2011). However, this does not allow the straightforward estimation of uncertainty in all types of signatures that is made possible by our method of generating multiple feasible realisations of rainfall and discharge time series.

6 Conclusions

This study investigated the effect of uncertainties in data and calculation methods on hydrological signatures. We present a widely-applicable method to evaluate signature uncertainty, and show results for two example catchments. The uncertainties were often large (i.e. typical intervals of ±10–40% relative uncertainty) and highly variable between signatures. It is therefore important to consider uncertainty when signatures are used for hydrological and ecohydrological analyses and modelling. Uncertainties of these magnitudes could change the conclusions of analyses such as cross-catchment comparisons or inferences about dominant processes.

Although we show that significant uncertainty can exist in hydrological signatures, we do not intend that this paper has a negative message. Consideration of uncertainty is equivalent to extracting the signal from noisy data, and not overestimating the information content in the data. As argued by Pappenberger and Beven (2006) and Juston et al. (2013), ignorance is not bliss when it comes to hydrological uncertainty; incorporation of uncertainty analysis leads to many advantages including more reliable and robust conclusions, reduction in predictive bias, and improved understanding. In particular, we hope that this paper encourages others to estimate data uncertainty in their catchments either individually or by reference to typical uncertainty magnitudes, to design diagnostic signatures and hypothesis testing techniques that are robust to data uncertainty, and to evaluate analysis results in the context of signature uncertainty.

Acknowledgements

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° 329762, by NIWA under Hazards Research Programme 1 (2014/15 SCI) and by Ministry of Business, Innovation and Employment, NZ, through contract C01X1006 Waterscape. We thank Alberto Viglione, Hoshin Gupta and Nataliya Le Vine for their constructive reviews which helped to improve this paper.

References

Ali, G., Oswald, C. J., Spence, C., Cammeraat, E. L. H., McGuire, K. J., Meixner, T., and Reaney, S. M.: Towards a unified threshold-based hydrological theory: necessary components and recurring challenges, Hydrological Processes, 27, 313-318, 2013.

Archer, D. and Newson, M.: The use of indices of flow variability in assessing the hydrological and instream habitat impacts of upland afforestation and drainage, J Hydrol, 268, 244-258, 2002.

Bengtsson, L. and Milloti, S.: Extreme storms in Malmo, Sweden, Hydrol Process, 24, 3462-3475, 2010.

Blazkova, S. and Beven, K. J.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, Water Resour Res, 45, W00B16, doi:10.1029/2007WR006726, 2009.

Bloeschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. H. G. (Eds.): Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales, Cambridge University Press, Cambridge, 2013.

Brown, J. D.: Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief, T I Brit Geogr, 29, 367-381, 2004.

Brutsaert, W. and Lopez, J. P.: Basin-scale geohydrologic drought flow features of riparian aquifers in the southern Great Plains, Water Resources Research, 34, 233-240, 1998.

Casper, M. C., Grigoryan, G., Gronz, O., Gutjahr, O., Heinemann, G., Ley, R., and Rock, A.: Analysis of projected hydrological behavior of catchments based on signature indices, Hydrol. Earth Syst. Sci., 16, 409-421, 2012.

Ciach, G. J.: Local random errors in tipping-bucket rain gauge measurements, Journal of Atmospheric and Oceanic Technology, 20, 752-759, 2003.

Clark, M., McMillan, H., Collins, D., Kavetski, D., and Woods, R.: Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, Hydrological Processes, 25, 523-543, 2011.

Clark, M. P., Rupp, D. E., Woods, R. A., Tromp-van Meerveld, H. J., Peters, N. E., and Freer, J. E.: Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale, Hydrological Processes, 23, 311-319, 2009.

Clausen, B. and Biggs, B. J. F.: Flow variables for ecological studies in temperate streams: groupings based on covariance, J Hydrol, 237, 184-197, 2000.

Coxon, G., Freer, J., Wagener, T., Odoni, A., and Clark, M.: Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrol. Process., 28, 6135–6150, doi:10.1002/hyp.10096, 2013.

Coxon, G. Freer, J. Westerberg, I.K., Wagener, T., Woods, R. and Smith, P.J. A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. Water Resour. Res., doi:10.1002/2014WR016532, 2015.

Eischeid, J. K., Baker, C. B., Karl, T. R., and Diaz, H. F.: The Quality-Control of Long-Term Climatological Data Using Objective Data-Analysis, Journal of Applied Meteorology, 34, 2787-2795, 1995.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol Earth Syst Sc, 17, 1893-1912, 2013.

Graham, C. B., van Verseveld, W., Barnard, H. R., and McDonnell, J. J.: Estimating the deep seepage component of the hillslope and catchment water balance within a measurement uncertainty framework, Hydrol Process, 24, 3631-3647, 2010a.

Graham, C. B., Woods, R. A., and McDonnell, J. J.: Hillslope threshold response to rainfall: (1) A field based forensic approach, Journal of Hydrology, 393, 65-76, 2010b.

Gupta, H. V., Wagener, T., and Liu, Y. Q.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol Process, 22, 3802-3813, 2008.

Gupta, H. V., and G. S. Nearing (2014), Debates—The future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science, Water Resour. Res., 50, 5351–5359, doi:10.1002/2013WR015096.

Gustard, A., Bullock, A., and Dixon, J. M.: Low flow estimation in the United Kingdom Institute of Hydrology, Wallingford, UK108, 88 pp., 1992.

Hall, F. R.: Base flow recessions - a review., Water Resources Research, 4, 973-983, 1968.

Hamilton, A. and Moore, R. D.: Quantifying Uncertainty in Streamflow Records, Canadian Water Resources Journal, 37, 3-21, 2012.

Harman, C. J., Sivapalan, M., and Kumar, P.: Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics, Water Resour. Res., 45, W09404, doi:10.1029/2008WR007392, 2009.

Harman, C. J., Troch, P. A., and Sivapalan, M.: Functional model of water balance variability at the catchment scale: 2. Elasticity of fast and slow runoff components to precipitation change in the continental United States, Water Resour. Res., 47, W02523, doi:10.1029/2010WR009656, 2011.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resour Res, 50, 7445–7469, 2014.

Jowett, I. G. and Duncan, M. J.: Flow Variability in New-Zealand Rivers and Its Relationship to in-Stream Habitat and Biota, New Zeal J Mar Fresh, 24, 305-317, 1990.

Juston, J., Jansson, P.-E., and Gustafsson, D.: Rating curve uncertainty and change detection in discharge time series: case study with 44-year historic data from the Nyangores River, Kenya, Hydrol Process, 28, 2509-2523, 2014.

Juston, J. M., Kauffeldt, A., Quesada Montano, B., Seibert, J., Beven, K. J., and Westerberg, I. K.: Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling, Hydrological Processes, 27, 1117-1122, 2013.

Kapangaziwiri, E., Hughes, D. A., and Wagener, T.: Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, Hydrolog Sci J, 57, 1000-1019, 2012.

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C. Y., and Westerberg, I. K.: Disinformative data in large-scale hydrological modelling, Hydrol Earth Syst Sc, 17, 2845-2857, 2013.

Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resources Research, 47, W11511, doi:10.1029/2011WR010748, 2011.

Kennard, M. J., Mackay, S. J., Pusey, B. J., Olden, J. D., and Marsh, N.: Quantifying Uncertainty in Estimation of Hydrologic Metrics for Ecohydrological Studies, River Res Appl, 26, 137-156, 2010.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resources Research, 45, W02429, doi:10.1029/2008WR006912, 2009.

Krueger, T., Freer, J., Quinton, J. N., Macleod, C. J. A., Bilotta, G. S., Brazier, R. E., Butler, P., and Haygarth, P. M.: Ensemble evaluation of hydrological model hypotheses, Water Resour. Res., 46, W07516, doi:10.1029/2009WR007845, 2010.

Li, H. Y. and Sivapalan, M.: Effect of spatial heterogeneity of runoff generation mechanisms on the scaling behavior of event runoff responses in a natural river basin, Water Resour. Res., 47, W00H08, doi:10.1029/2010WR009712, 2011.

McDonnell, J. J. and Woods, R.: On the need for catchment classification, Journal of Hydrology, 299, 2-3, 2004.

McMillan, H., Clark, M., Bowden, W. B., Duncan, M. J., and Woods, R.: Hydrological field data from a modeller's perspective. Part 1: Diagnostic tests for model structure, Hydrological Processes, 25, 511-522, 2011.

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., and Clark, M.: Impacts of uncertain flow data on rainfall-runoff model calibration and discharge predictions, Hydrological Processes, 24, 1270-1284   2010.

McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M. P., and Rupp, D. E.: Spatial variability of hydrological processes and model structure diagnostics in a 50 km² catchment, Hydrol. Process., 28, 4896–4913, doi:10.1002/hyp.9988, 2014.

McMillan, H. and Westerberg, I. K.: Rating curve estimation under epistemic uncertainty, Hydrological Processes, 29(7), 1873-1882, 2015

McMillan, H. K., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol Process, 26, 4078-4111, 2012.

Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resour. Res., 43, W05434, doi:10.1029/2006WR005184, 2007.

Moore, R. J., Jones, D. A., Cox, D. R., and Isham, V. S.: Design of the HYREX raingauge network, Hydrol. Earth Syst. Sci., 4, 521–530, doi:10.5194/hess-4-521-2000, 2000.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Res Appl, 19, 101-121, 2003.

Pappenberger, F. and Beven, K. J.: Ignorance is bliss: Or seven reasons not to use uncertainty analysis, Water Resour. Res., 42, W05302 doi:10.1029/2005WR004820, 2006.

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J. B., Pfister, L., and Fraipont de, P.: Influence of uncertain boundary conditions and model structure on flood inundation predictions, Adv Water Resour, 29, 1430-1449, 2006.

Pechlivanidis, I. G., Jackson, B. M., McMillan, H. K., and Gupta, H. V.: Using an informational entropy-based metric as a diagnostic of flow duration to drive model parameter identification, Global Nest Journal, 14, 325-334, 2012.

Penna, D., Tromp-van Meerveld, H. J., Gobbi, A., Borga, M., and Dalla Fontana, G.: The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment, Hydrol. Earth Syst. Sci., 15, 689-702, 2011.

Pokhrel, P., Yilmaz, K. K., and Gupta, H. V.: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, Journal of Hydrology, 418, 49-60, 2012.

Refsgaard, J. C. and Knudsen, J.: Operational validation and intercomparison of different types of hydrological models, Water Resour Res, 32, 2189-2202, 1996.

Rupp, D. E. and Selker, J. S.: Information, artifacts, and noise in dQ/dt-Q recession analysis, Advances in Water Resources, 29, 154-160, 2006.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol Earth Syst Sc, 15, 2895-2911, 2011.

Seibert, J. and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resour. Res., 38, 1241, doi:10.1029/2001WR000978, 2002.

Shaw, S. B. and Riha, S. J.: Examining individual recession events instead of a data cloud: Using a modified interpretation of dQ/dt-Q streamflow recession in glaciated watersheds to better inform models of low flow, Journal of Hydrology, 434, 46-54, 2012.

Sieck, L. C., Burges, S. J., and Steiner, M.: Challenges in obtaining reliable measurements of point rainfall, Water Resour. Res., 43, W01420 doi:10.1029/2005wr004519, 2007.

Sivakumar, B., Singh, V., Berndtsson, R., and Khan, S.: Catchment Classification Framework in Hydrology: Challenges and Directions, Journal of Hydrologic Engineering, 10.1061/(ASCE)HE.1943-5584.0000837 , A4014002, 2013.

Smith P, Beven KJ, Tawn JA.: Informal likelihood measures in model assessment: theoretic development and investigation. Advances in Water Resources 31(8): 1087–1100, 2008.

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., and Tallaksen, L. M.: Comparison of hydrological model structures based on recession and low flow simulations, Hydrol. Earth Syst. Sci., 15, 3447-3459, 2011.

Stoelzle, M., Stahl, K., and Weiler, M.: Are streamflow recession characteristics really characteristic?, Hydrol Earth Syst Sc, 17, 817-828, 2013.

Sugawara, M.: Automatic calibration of the tank model, Hydrological Sciences Bulletin, 24, 375-388, 1979.

Tallaksen, L. M.: A review of baseflow recession analysis, Journal of Hydrology, 165, 349-370, 1995.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope, Water Resour. Res., 42, W02410, doi:10.1029/2004WR003778, 2006a.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, Water Resour. Res., 42, W02411, doi:10.1029/2004WR003800, 2006b.

Villarini, G. and Krajewski, W. F.: Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges, Advances in Water Resources, 31, 1015-1023, 2008.

Viney, N. R. and Bates, B. C.: It never rains on Sunday: The prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set, Int J Climatol, 24, 1171-1192, 2004.

Wagener, T., Sivapalan, M., Troch, P. A., and Woods, R.: Catchment classification and hydrologic similarity, Geography compass, 1, 901-931, 2007.

Westerberg, I. K., Gong, L., Beven, K., Seibert, J., Semedo, A., Xu, C. Y., and Halldin, S.: Regional water balance modelling using flow-duration curves with observational uncertainties, Hydrol Earth Syst Sc, 18, 2993-3013, 2014.

Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C. Y.: Calibration of hydrological models using flow-duration curves, Hydrol Earth Syst Sc, 15, 2205-2227, 2011.

Westerberg, I. K., Walther, A., Guerrero, J.-L., Coello, Z., Halldin, S., Xu, C. Y., Chen, D., and Lundin, L.-C.: Precipitation data in a mountainous catchment in Honduras: quality assessment and spatiotemporal characteristics, Journal of Theoretical and Applied Climatology, 101, 381-396, 2010.

Winsemius, H. C., Schaefli, B., Montanari, A., and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, Water Resour. Res., 45, W12422, doi:10.1029/2009wr007706, 2009.

WMO: Guide to hydrological practices, vol I.: hydrology— from measurement to hydrological information World Meteorological Organization, Geneva168, 2008.

Wood, S. J., Jones, D. A., and Moore, R. J.: Accuracy of rainfall measurement for scales of hydrological interest, Hydrol Earth Syst Sc, 4, 531-543, 2000.

Woods, R. A., Grayson, R. B., Western, A. W., Duncan, M. J., Wilson, D. J., Young, R. I., Ibbitt, R. P., Henderson, R. D., and McMahon, T. A.: Experimental Design and Initial Results from the Mahurangi River Variability Experiment: MARVEX. In: Observations and Modelling of Land Surface Hydrological Processes, Lakshmi, V., Albertson, J. D., and Schaake, J. (Eds.), Water Resources Monographs, American Geophysical Union, Washington D.C., 201–213, 2001.

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Adv Water Resour, 30, 1756-1774, 2007.

Zhang, Y., Vaze, J., Chiew, F. H. S., Teng, J., and Li, M.: Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall-runoff modelling, Journal of Hydrology, 517, 936-948, 2014.

Tables

Table 1. Sources of uncertainty considered in this study, and the methods used for estimation.

Variable/Signature	**Uncertainty sources**	**Estimation method**	**Reference if applicable**
Rainfall	Point uncertainty	Normal distribution with σ a function of rain rate.	Ciach (2003)
	Interpolation uncertainty	Subsampling from a dense network of rain gauges	
	Equipment malfunction	Rainfall data with/without QC	Wood et al. (2000)
Flow	Discharge uncertainty in gaugings	Analysis of stations with stable ratings	Coxon et al. (2015)
	Stage uncertainty in gaugings	Uniform distribution ±5 mm	McMillan et al. (2012)
	Rating-curve uncertainty	Voting Point likelihood method	McMillan and Westerberg (2015)
Recession analysis	Flow data time step	Tested hourly vs daily	
	Seasonality of response	Tested using all data or split by season	Shaw and Riha (2012)
Rainfall-runoff threshold	Effects of baseflow	Tested with/without baseflow separation	Gustard et al. (1992)
	Rainfall event definition	Tested with/without inclusion of smaller events	

Table 2~~1~~. Basic rainfall–runoff signatures included in the study. All signatures are calculated on hourly data unless otherwise specified.

Signature	Name	Description	Unit
Flow distribution			
Q_{MEAN}	Mean flow	Mean flow for the analysis period	mm h^{-1}
$Q_{0.01}$,$Q_{0.1}$,Q_{1},Q_{5}, Q_{50},Q_{85},Q_{95},Q_{99}	Flow percentiles	Low and high flow exceedance percentiles from the FDC	mm h^{-1}
Event frequency and duration			
Q_{HF}	High flow event frequency	Average number of daily high flow events per year, with a threshold of 9 times the median daily flow (Clausen and Biggs, 2000)	yr^{-1}
Q_{HD}	High flow event duration	Average duration of daily flow events higher than 9 times the median daily flow (Clausen and Biggs, 2000)	days
Q_{LF}	Low flow event frequency	Average number of daily low flow events per year, with a threshold of 0.2 times the mean daily flow (Olden and Poff, 2003, they used a 5% threshold)	yr^{-1}
Q_{LD}	Low flow event duration	Average duration of daily flow events lower than 0.2 times the mean daily flow (see Q_{LF})	days
Flow dynamics			
BFI	Base Flow Index	Contribution of baseflow to total streamflow, calculated from daily flows using the Flood Estimation	-

		Handbook method (Gustard et al., 1992)	
S_{FDC}	Slope of the normalised FDC	Slope of the FDC between 33% and 66% exceedance values of streamflow normalised by its mean (Yadav et al., 2007)	-
Q_{CV}	Overall flow variability	Coefficient of variation in streamflow, i.e. standard deviation divided by mean flow (Clausen and Biggs, 2000; Jowett and Duncan, 1990)	-
Q_{LV}	Low flow variability	Mean of annual minimum flow divided by the median flow (Jowett and Duncan, 1990).	-
Q_{HV}	High flow variability	Mean of annual maximum flow divided by the median flow (Jowett and Duncan, 1990).	-
Q_{AC}	Flow autocorrelation	Autocorrelation for 1 day (24 hours). Used by (Euser et al., 2013) and (Winsemius et al., 2009).	-
Rainfall-runoff			
RR	Total runoff ratio	Total runoff divided by total precipitation	-
Rainfall			
P_{MA}	Mean annual precipitation	Mean annual catchment average precipitation	mm yr^{-1}
P_{STD}	Standard deviation	Standard deviation of catchment	mm h^{-1}

	of hourly precipitation	average precipitation

Table 3 Dominant uncertainty sources and uncertainty characteristics

Signature type		**Catchment[1]**	**Dominant uncertainty source**	**Uncertainty characteristics**			
				Half-width of 5–95 percentile range (%)	Mean (=bias) (%)	Std. (%)	Skew-ness (-)
Flow distribution	Average flow conditions (Q_{MEAN})	M	Rating-curve uncertainty	11.1	-0.4	6.8	0.32
		B	Rating-curve uncertainty	12.7	-2.4	7.7	-0.03
	Low flow percentiles (Q_{95})	M	Discharge gauging uncertainty	23.8	-1.2	14.6	0.47
		B	Rating-curve uncertainty	39.5	-1.1	23.8	0.45
	High flow percentiles ($Q_{0.1}$)	M	Rating-curve uncertainty	22.8	-8.3	16.6	1.54
		B	Rating-curve uncertainty	19.6	0.0	12.0	0.13
Events	Event frequency and duration (Q_{HD})	M	Threshold value, which depends on rating-curve uncertainty	6.9	2.3	3.3	1.30
		B	Threshold value, which depends on rating-curve uncertainty	21.6	-5.1	13.1	0.57
Flow dynamics	Base Flow Index (BFI)	M	Rating-curve uncertainty	11.6	3.4	7.1	-0.11
		B	Rating-curve uncertainty	8.5	-2.3	5.1	-0.19
	Slope of Flow Duration	M	Rating-curve breakpoint location	28.8	16.9	17.4	0.46

	Curve (S_{FDC})	B	Rating-curve uncertainty	6.0	-3.2	3.7	-0.18
	Variability of extreme flows (Q_{HV})	M	Rating-curve uncertainty	41.9	-1.0	30.4	2.30
		B	Rating-curve uncertainty	37.0	6.5	23.0	0.75
	Recession analysis (b hourly)	M	Calculation time step	9.9	-3.1	6.3	0.38
		B	Rating-curve uncertainty	14.9	5.1	8.9	0.72
Rainfall-runoff	Total runoff ratio (RR[2])	M	Rating-curve uncertainty	14.6	-0.3	9.0	0.26
		B	Rating-curve uncertainty	13.3	-2.0	8.1	0.02
	Rainfall-runoff threshold (threshold location[3])	M	Rainfall interpolation uncertainty	17.3	16.3	17.2	5.88
		B	-	-	-	-	-
Rainfall	Mean annual precipitation (P_{MA}[2])	M	Interpolation uncertainty	10.0	0.3	5.7	0.22
		B	Interpolation uncertainty. (Equipment malfunction)	4.6	-0.4	2.7	0.34
	Standard deviation of precipitation (P_{STD}[2])	M	Interpolation uncertainty	8.0	9.5	4.4	1.55
		B	Interpolation uncertainty. (Equipment malfunction)	4.9	4.6	2.9	0.67

[1] M = Mahurangi, B = Brue

[2] These signatures were calculated using 1 gauge/45 km^2 and including point error

[3] This signature was calculated for the total uncertainty scenario in Fig. 10

Figures

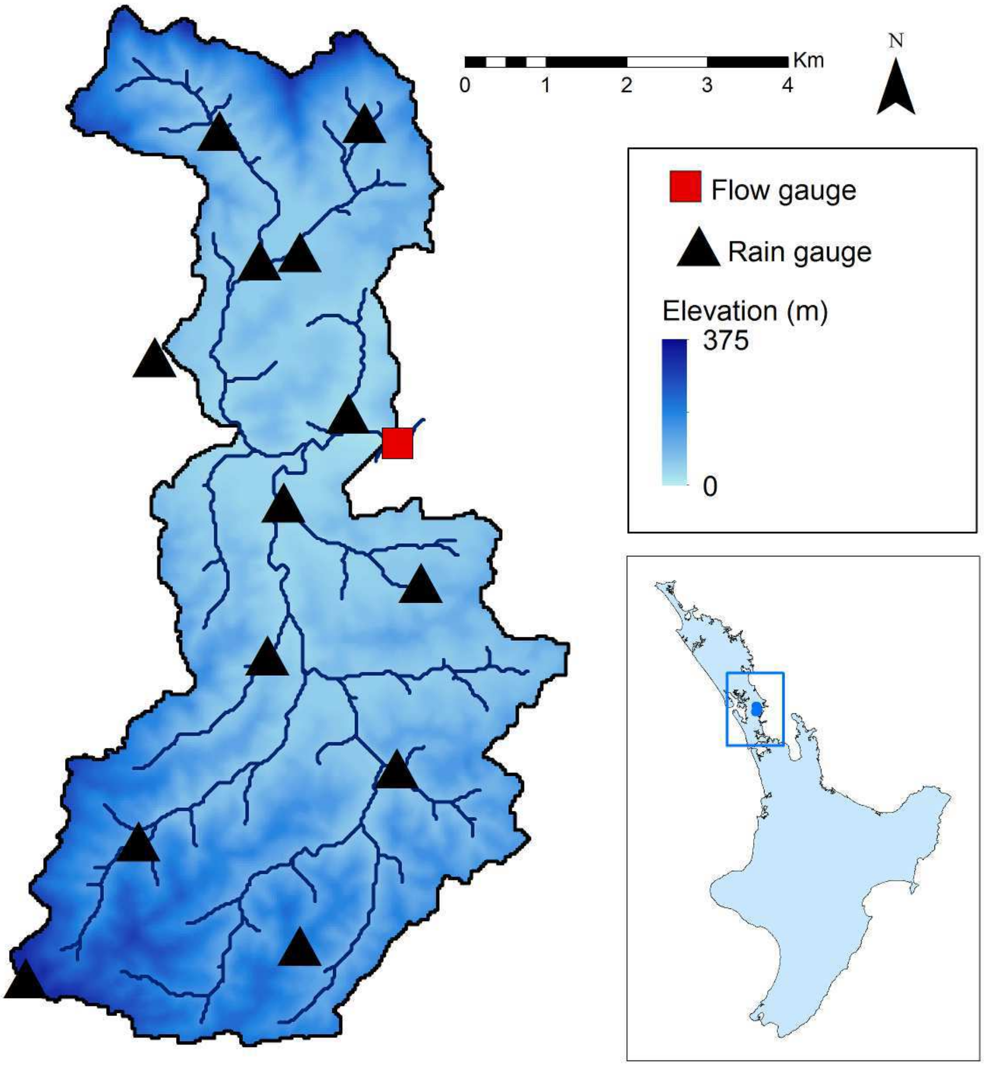

Fig. 1 The Mahurangi catchment in New Zealand and the location of the rain gauges and the outlet flow gauge.

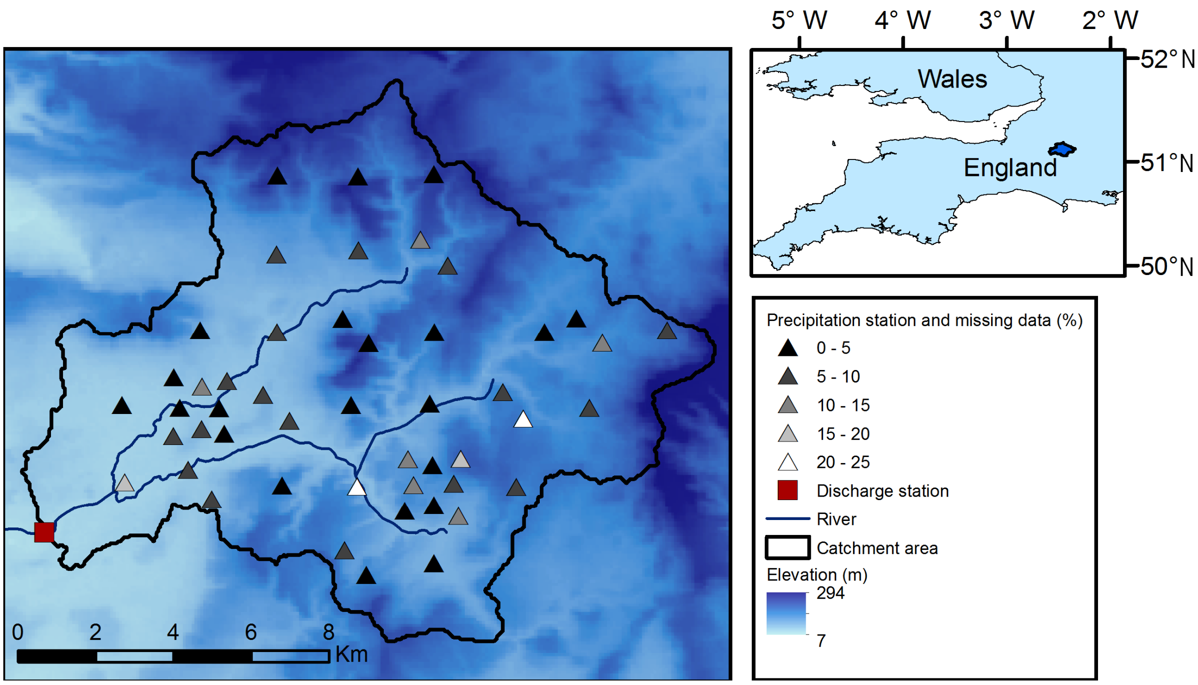

Fig. 2 The Brue catchment in south-west England, and the location of the precipitation and discharge stations. The percent of missing values after quality control is given for each rain gauge.

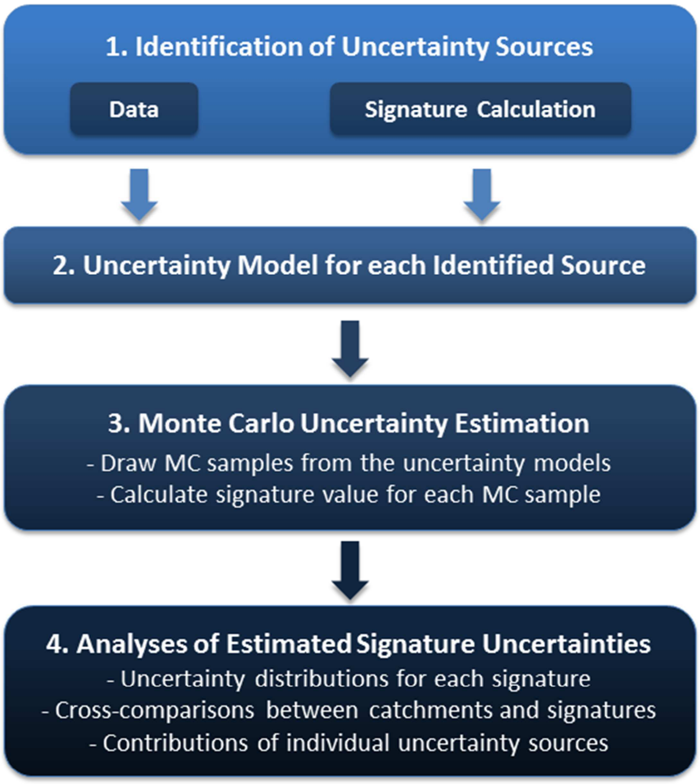

Fig. 3 Schematic description of the method used for estimation of signature uncertainty

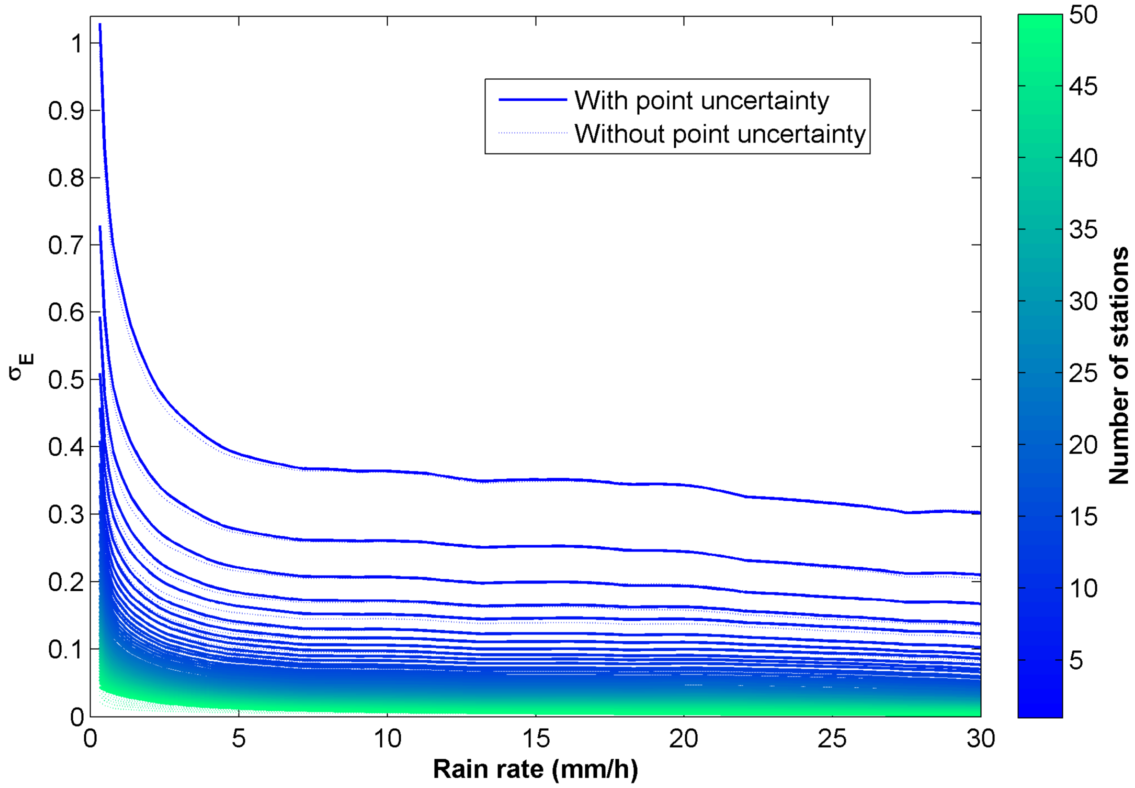

Fig. 4. Standard deviation of the rainfall error as a function of rain rate for different numbers of subsampled stations for 1000 Monte Carlo realisation for the Brue catchment, with and without point uncertainty.

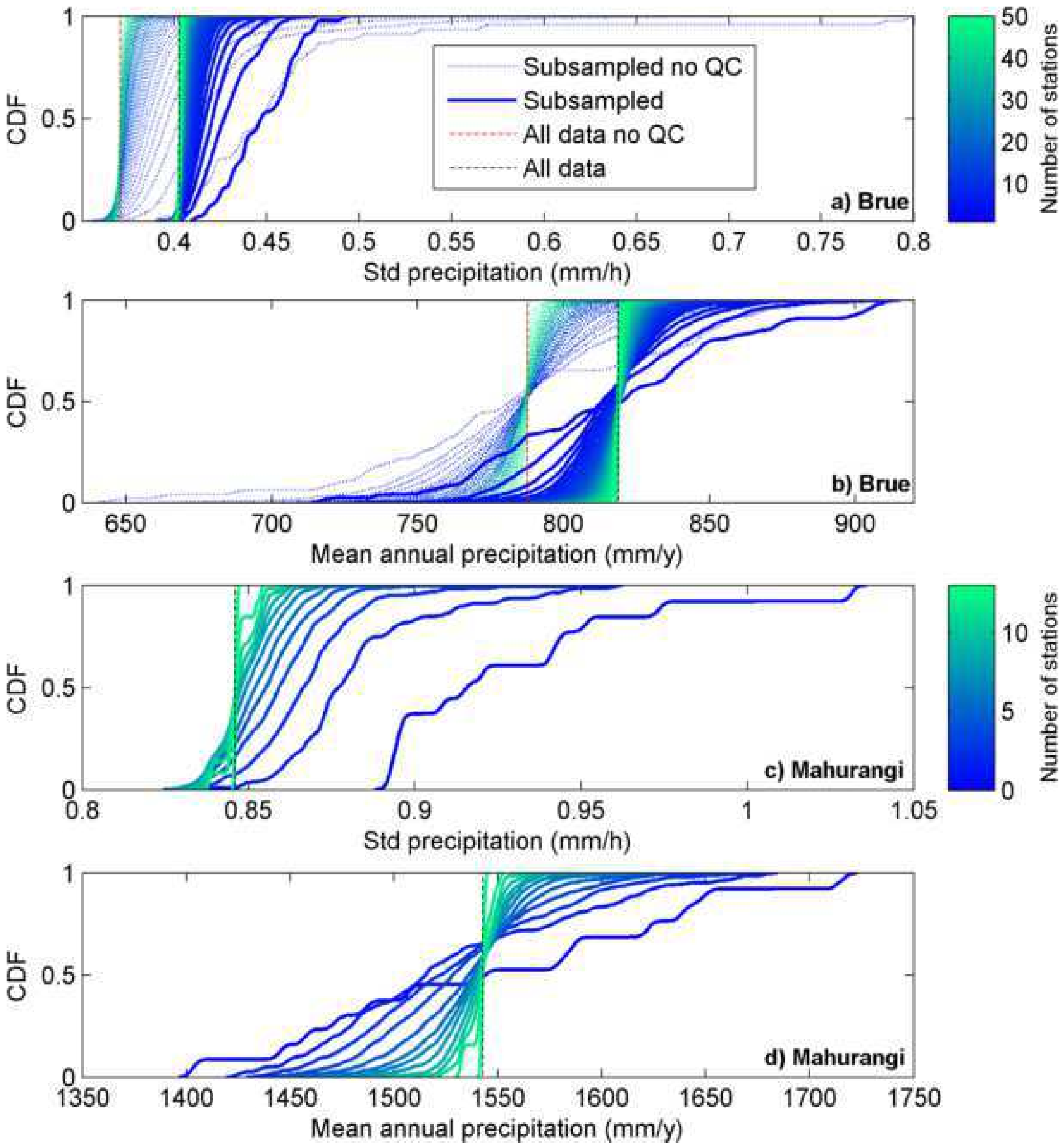

Fig. 5 a) and c) standard deviation of hourly precipitation and, b) and d), mean annual precipitation for different numbers of subsampled stations. For the Mahurangi results are shown for the period without missing discharge values. Point measurement uncertainty was included and we used 4000 Monte Carlo realisations.

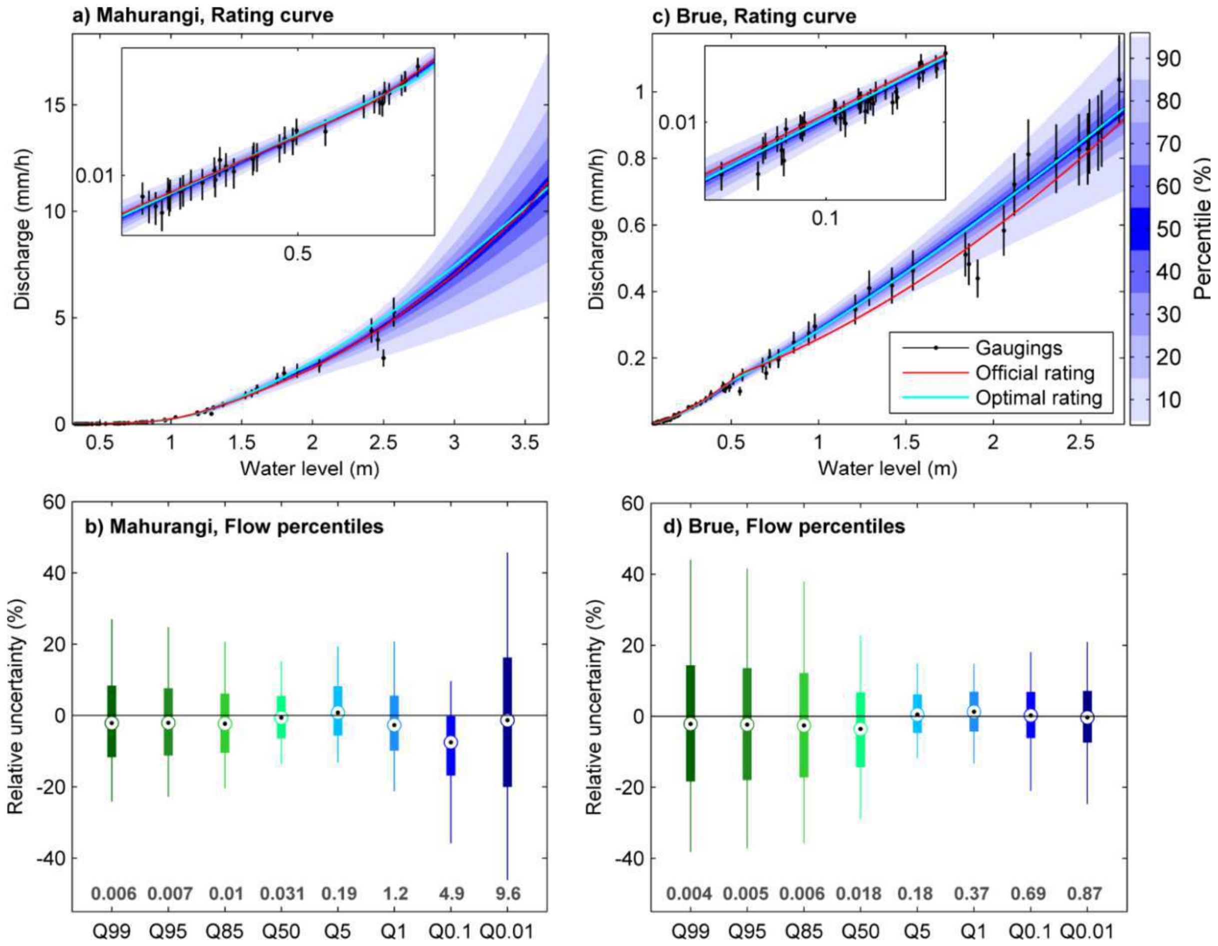

Fig. 6 Estimated rating curve uncertainty and uncertainty in flow percentiles for the Mahurangi (a and b) and Brue (c and d) catchments. Uncertainties are calculated relative to the optimal rating curve from the MCMC. For Brue the official rating curve is dissimilar to the optimal MCMC rating curve because it was calculated for a longer gauging dataset starting in the 1960's, with considerably more variability. The rating curve is shown in linear space, with an inset plot in log space for the low-flow range. The flow percentiles for the official rating are given as hourly averages in mm h^{-1} in the bottom of the (b and d) figures. The boxplot whiskers extend to the 5 and 95 percentiles, and the box covers the interquartile range.

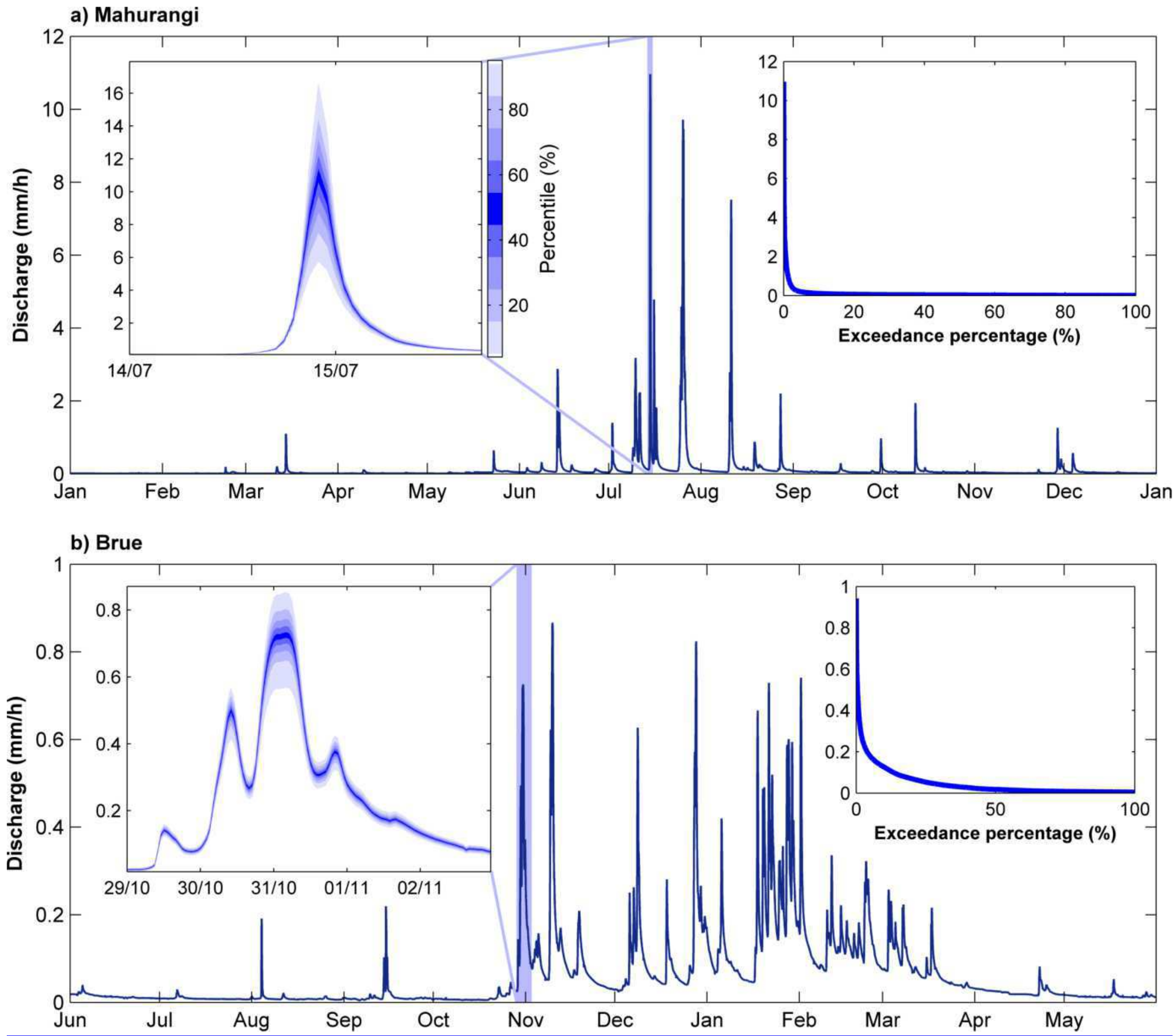

Fig. 7 Discharge calculated using the optimal rating curve for 1998 for Mahurangi (a) and for 1994–1995 for Brue (b). The left inset plots show the discharge time series uncertainty distribution at an hourly scale for a peak flow event in each catchment. The right inset plots show the flow-duration curves for the full time series for each catchment. The y-axis variable and unit is discharge in mm/h in all plots.

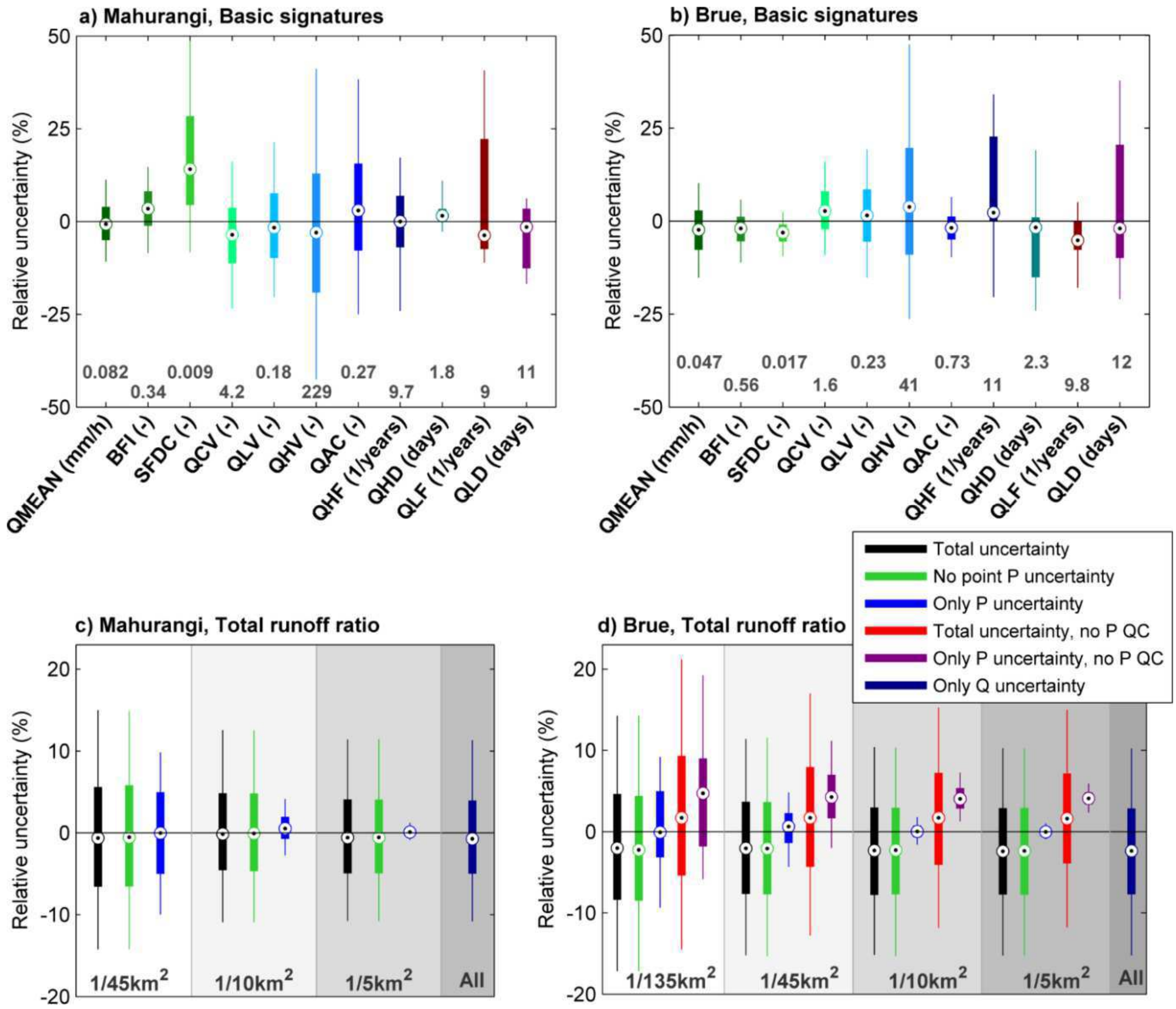

Fig. 8~~7~~ Relative uncertainty in basic signatures as a percentage of the signature values calculated with the optimal rating curve from the MCMC. The boxplot whiskers extend to the 5 and 95 percentiles, and the box covers the interquartile range.

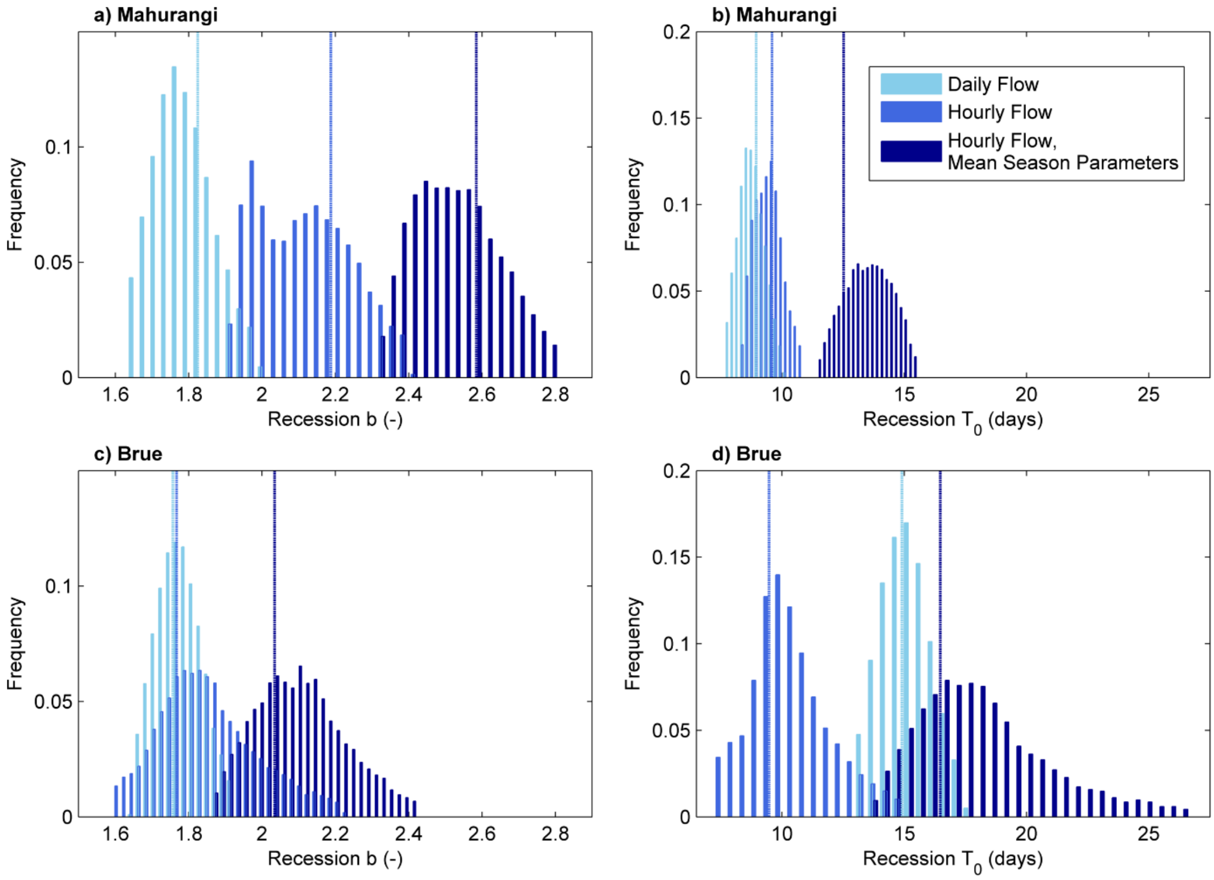

Fig. 9~~8~~ Histograms of recession parameter distributions, where parameters are calculated using (1) daily flow data, (2) hourly flow data, and (3) hourly flow data where recession parameters are calculated per season and then averaged. Dotted lines show the parameter values from the optimal MCMC rating curve. Distributions are truncated at the 2.5 and 97.5 percentiles.

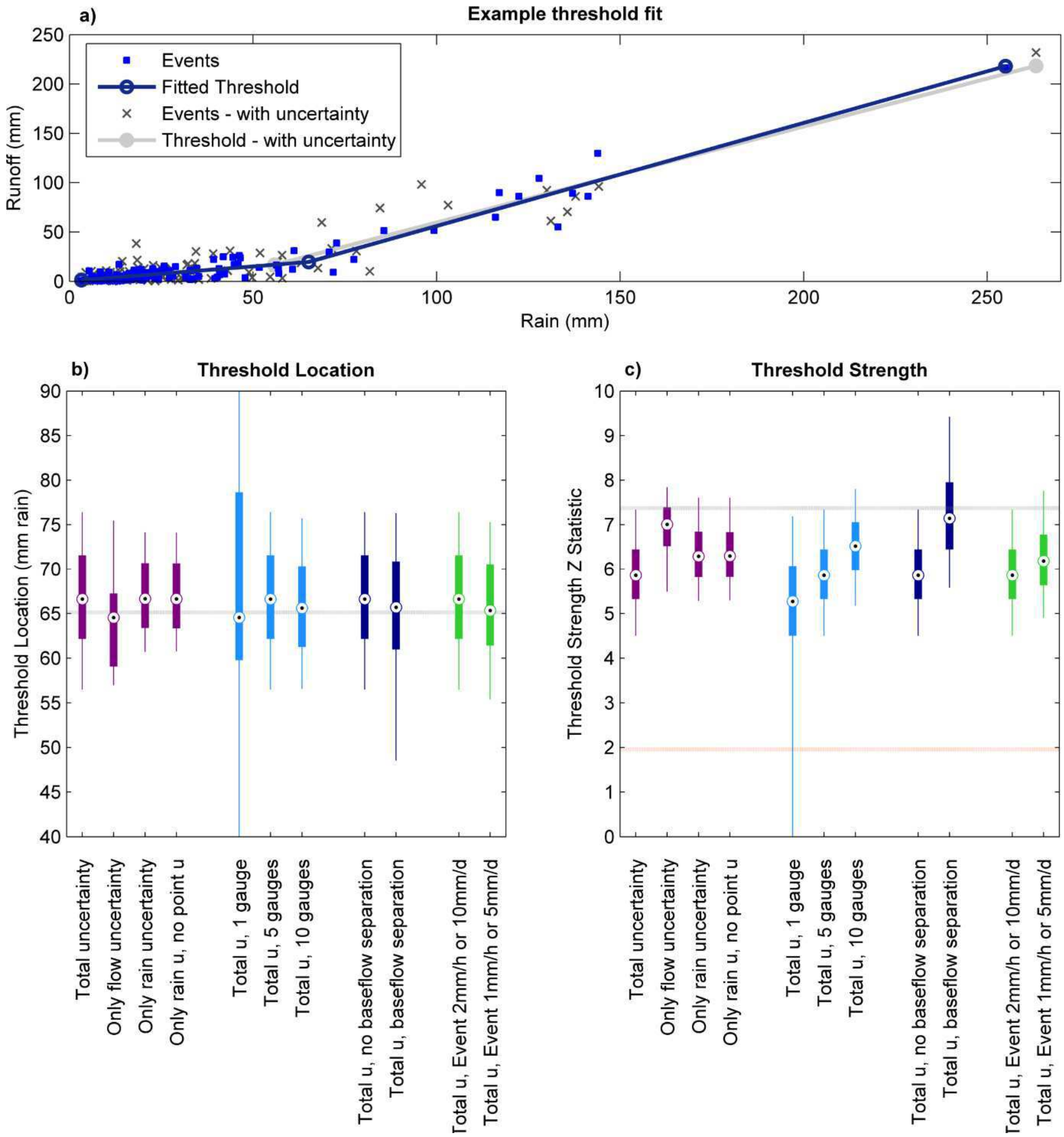

Fig. 10~~9~~ a) example of the threshold fitting procedure without (blue) and with (grey, one raingauge scenario) uncertainty. Box plots of b) threshold location and c) threshold strength in the Mahurangi catchment, under different data and subjective uncertainty scenarios. Horizontal grey lines show baseline signature values from the optimal rating curve and precipitation data. The orange line in Fig. 9c shows the value above which the change in slope of the rainfall-runoff relationship is significant at the 5% level. Boxplot whiskers for the uncertainty distribution in the 1 raingauge scenario are truncated for clarity. The total uncertainty scenario used 1 raingauge per 10 km^2.