Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality

Journal:	*Hydrological Processes*
Manuscript ID:	HYP-11-0442.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	McMillan, Hilary; National Institute of Water and Atmospheric Research, - Krueger, Tobias; University of East Anglia, School of Environmental Sciences Freer, Jim; University of Bristol, School of Geographical Sciences
Keywords:	Observational uncertainty, Data uncertainty, Hydrology, Error distributions, Hydrometric data, Water quality data

Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality

Hilary McMillan[1], Tobias Krueger[2], Jim Freer[3]

[1]National Institute of Water and Atmospheric Research, Christchurch, New Zealand

[2]School of Environmental Sciences, University of East Anglia, Norwich, UK

[3]School of Geographical Sciences, University of Bristol, Bristol, UK

Abstract

This review and commentary sets out the need for authoritative and concise information on the expected error distributions and magnitudes in observational data. We discuss the necessary components of a benchmark of dominant data uncertainties, and the recent developments in hydrology which increase the need for such guidance. We initiate the creation of a catalogue of accessible information on characteristics of data uncertainty for the key hydrological variables of rainfall, river discharge and water quality (suspended solids, phosphorus, nitrogen). This includes demonstration of how uncertainties can be quantified, summarising current knowledge and the standard quantitative results available. In particular, synthesis of results from multiple studies allows conclusions to be drawn on factors which control the magnitude of data uncertainty, and hence improves provision of prior guidance on those uncertainties. Rainfall uncertainties were found to be driven by spatial scale, while river discharge uncertainty was dominated by flow condition and gauging method. Water quality variables presented a more complex picture with many component errors. For all variables it was easy to find examples where relative error magnitudes exceeded 40%. We consider how data uncertainties impact on the interpretation of catchment dynamics, model regionalisation and model evaluation. In closing the review, we make recommendations for future research priorities in quantifying data uncertainty, and highlight the need for an improved 'culture of engagement' with observational uncertainties.

1 Introduction

Considerable attention has been given in recent literature to the challenges of the hydrological modelling process. A key question is how our community should move forward to improve understanding of hydrological systems and simulation models, in the light of uncertain observed data. Notable opinion papers that tackle these issues have been published regarding general approaches to uncertainty (Pappenberger and Beven, 2006); uncertainty frameworks (Refsgaard et al., 2007); equifinality (Beven, 2006); model diagnostics (Gupta et al., 2008); uncertainty frameworks for ungauged basins (Wagener et al., 2006); and uncertainty assessments in water quality modelling

(Beck, 1987). The modelling community has also engaged in substantial discussion and counter discussion about the merits for making strong and weak judgements about the nature of data, parameter and structural errors (Beven et al., 2007; Beven, 2009; Beven et al., 2008; Beven et al., 2011; Mantovan and Todini, 2006; Mantovan et al., 2007; Montanari, 2005; Schoups and Vrugt, 2010; Stedinger et al., 2008; Vrugt et al., 2009a; 2009b). Quantifying the expected data errors, or at least developing an informed and rational framework for approximating them, is arguably prerequisite for understanding the other error sources. Yet, despite extensive discussions, there remains a lack of authoritative and concise information on the expected error distributions and magnitudes in observational data.

We suggest in this review and commentary that we must improve our benchmarking of the dominant uncertainties in observational data. This is necessary for the hydrological community to develop effective guidance on uncertainty frameworks for modelling objectives and hence credible measures of model performance. Benchmarking, expressed in these terms, necessarily includes understanding the effectiveness of data to characterise hydrological processes under spatial and temporal heterogeneity, before any modelling is conducted. This is a holistic view of the information content of data, recognising that different monitoring technologies characterise to a greater or lesser extent the quantity of interest irrespective of the implied precision and accuracy of instrumentation. Some observed responses may be subject to commensurability error; i.e. they are not well represented at the relevant temporal and spatial resolution. This may over-emphasise a belief of process characterisation at a point or quasi-point scale that is not appropriate to the actual variability of the process occurring over the conceptualised control volume and/or period. Ultimately, these issues could misguide inferences in space and time and may impact on our ability to develop appropriate theory and conceptualisations. Benchmarking of course also includes understanding the precision of instruments and calibrations and inadequacies in transforming measurements (e.g. stage to discharge, concentration to load). A suggested benchmark may necessarily be crude initially, due to limitations of current measurement technologies to quantify all error sources precisely, but we strongly feel that some guidance is required to bridge the current disconnect between how we use data to evaluate models and what the possible errors and information content might be in such data.

We argue that several recent developments in hydrology and water quality research increase the need for guidance and a framework for understanding the potential for data errors. There is a growing need to interrogate a diversity of data sources that elucidate different processes and states in catchment behaviour across multiple time and space scales. This need is driven by the multidisciplinary nature of the scientific problems we are asked to solve, e.g. understanding environmental change impacts on the ecological status of river systems, understanding effectiveness of policy interventions; however, data uncertainties can propagate into such environmental management investigations (Mahmoud et al., 2009). Diverse data types are also increasingly co-interrogated to test coherence of internal catchment states and evaluate simulated flow dynamics on the quest for models which produce “the right answers for the right reasons” (Kirchner, 2006) and hence provide more robust predictions under changes in climate, land use or land management. The hydrological community is embracing the potential of additional data to improve model structure; but we must recognise that diverse data types introduce diverse error characteristics and respect those when developing evaluation methods (see Section 6.1 for further discussion). Another development is in recognition of the value of comparative hydrology and new approaches to catchment similarity and regionalisation, both of which will suffer inevitable bias if place-specific observational uncertainty is not properly included. Finally, we recognise the creation of 'virtual observatories' which are improving scientists' ability to discover and capture cross-nation data sources (e.g. http://www.evo-uk.org). Virtual observatories are driving the development of various data standards (e.g. Tarboton et al., 2008) which need to further evolve to increase the awareness of data quality in the metadata. We revisit the effects of uncertainty on studies incorporating multi-response data and on regionalisation studies in Sections 6.1 and 6.2, respectively.

There is no doubt that there is already a considerable wealth of information to start the process of providing 'frames of reference' to possible errors in hydrologically important information. Much is to be gained from a catalogue of this information in accessible form, which we aim to provide here. We set out the important characteristics of data uncertainty for key hydrological variables and how they can be quantified, summarising current knowledge and the standard quantitative results available to the practitioner. We hope to provide guidance which encourages and facilitates the important step of including estimates of measurement error and its effects within hydrological studies. At the very least we aim to start a dialogue that considers how the community can improve our quantification of data errors and how current accepted practice may sometimes bias our ability to understand observed or simulated catchment behaviour. This paper fully accepts that some aspects of data error may be problematic or impossible to define (e.g. one cannot estimate the spatial variability of a particular rainfall field from a single rain gauge), but this must not absolve inaction. We also identify future research priorities for improving the data uncertainty estimates listed herein. In closing the review, we discuss our recommendations and highlight the need for an improved 'culture of engagement' with observational uncertainties.

The scope of the paper is necessarily bounded. We focus our attention on uncertainty in measurements of rainfall (Section 3), river discharge (Section 4) and suspended solids, phosphorus and nitrogen concentrations (Section 5). However, the discussions are equally relevant to many other types of hydrological measurement uncertainty, such as in measurements of evapotranspiration (Llasat and Snyder, 1998), snow (Goodison et al., 1998), hydrogeological quantities including hydraulic conductivity (Nilsson et al., 2007), water table depth (Freer et al., 2004), soil physical and chemical properties (Owens et al., 2008; van der Keur and Iversen, 2006), topography (Wechsler, 2007), and land use (Castilla and Hay, 2007, in a remote sensing context). Our focus is to benchmark and explore uncertainties in data, not to consider how methods might be deployed for reducing uncertainty. There is also a large body of literature on impact studies, analysing the effects of uncertainty on simulation models, which we do not review here, but for which a clear inventory of uncertainty magnitudes is critical. Finally, we accept that we may have missed additional papers that have explored data uncertainty issues, therefore in Section 7.2 we encourage authors to report such oversights through the Experimental Hydrology Wiki to further improve our collective understanding of observational uncertainties

2 Uncertainty characterisation

Uncertainty in hydrological systems comes in many guises and there is a need to use clear terminology (Montanari, 2007). Di Baldassarre and Montanari (2009) summarise three primary sources in the context of hydrological modelling: (a) uncertainty in observations, (b) parameter uncertainty, and (c) model structural uncertainty. Our review clearly considers the first of these sources as a precursor to understanding the remaining two. Bogardi and Kundzewicz (1996) distinguish pre-hydrological (e.g. meteorology) and post-hydrological (e.g. social, psychological and institutional) sources of uncertainty. Klemeš (1996) provides a typically robust view that the types of uncertainty considered may be governed by social or political factors, and draws attention to the 'unknown unknowns': uncertainty sources which are unidentified until a hydrological event or disaster brings them to our attention. This echoes the division between *Natural Uncertainty* (also termed aleatory uncertainty, inherent variability, type-A uncertainty) and *Epistemic Uncertainty* (also lack-of-knowledge uncertainty, ignorance, type-B uncertainty) (Merz and Thieken, 2005).

Natural uncertainties can be treated formally by statistical probabilistic methods, although it can be difficult to identify an appropriate statistical model. Epistemic uncertainties imply that the nature of the uncertainty is not known precisely, and formal statistical methods can provide only an approximation to these uncertainties by treating them as if they were aleatory. In a hydrological data context, epistemic uncertainties may relate to spatial heterogeneity (e.g. in rainfall or evapotranspiration), transformation of variables (e.g. snow water equivalent, discharge rating curves) or lack of knowledge regarding boundary conditions (e.g. losses to deep groundwater) (Beven *et al.*, 2011). They may be non-stationary, e.g. vegetation growth at rain or flow gauge sites. Other sources can include malfunctioning measurement equipment and human-induced measurement errors (Viney and Bates, 2004; 'spurious errors': Herschy, 1998). Epistemic errors can occur during data management, storage or post-processing which may be largely undocumented. Recent work has emphasised that epistemic uncertainty can change the information content of observed data to the extent that some observations may be regarded as 'disinformation' where they contradict physical laws such as the water balance (Beven and Westerberg, 2011; Beven et al., 2011). We do however note that such hydrological outliers are often indicative of unmodelled processes, variability or boundary conditions (Andréassian et al., 2010). Although epistemic uncertainty is inherently difficult to quantify, it is important to avoid the tendency to ignore sources of uncertainty which are hard to measure.

The metrics by which observational uncertainty is quantified will depend on the types and sources of uncertainty, how well they are understood, and the estimation and reporting techniques being used. Uncertainties may be modelled as, in the order of increasing information content: mean relative/absolute error (which may include components of systematic bias and random error), upper and lower bounds, fuzzy membership functions or full probability distributions; or errors may simply be divided as acceptable/unacceptable (Bulygina et al., 2009). Error models may additionally contain information on the autocorrelation and heteroscedastic properties of the uncertainty. In our tabulation of uncertainty estimates we have made clear the metrics being used as far as this information was available from the original studies, as knowledge of these is needed to allow collation and comparison.

A thoughtful choice of uncertainty model is essential. Care should be taken that the choice is not biased by the availability of models in uncertainty propagation tools (e.g. Data Uncertainty Engine (DUE), Brown and Heuvelink, 2007). Assuming a probability distribution based on insufficient evidence will give a false sense of certainty of 2nd order ('certainty about uncertainty'; Brown, 2004). Conversely, reporting uncertainty via confidence bounds may miss later opportunities, e.g. for results to be incorporated into decision theoretic frameworks which require full uncertainty distributions. Evidence theory provides a compelling overarching framework (Hall, 2003). In some cases a sensitivity analysis exploring various *hypothetical* levels of uncertainty (e.g. Info-Gap theory, Ben-Haim, 2006) may be all that is justifiable, and indeed required. Brown (2004) suggests that the choice of complexity of uncertainty analysis may also be influenced by the level of risk associated with poor decisions and the resources availability for making those decisions (time, money, expertise).

In most applications there are many contributing sources of observational uncertainty: the method by which they are combined is therefore also important. Previous authors have typically used the root-mean-square (RMS) approach (e.g. Harmel et al., 2006; Sauer and Meyer, 1992) which assumes errors are non-additive; or may have varied the error structure according to uncertainty type (e.g. additive, absolute additive, RMS; Di Baldassarre and Montanari, 2009). It is also common for total uncertainty to be estimated by numerical simulation (e.g. Di Baldassarre and Montanari, 2009; McMillan et al., 2010). These methods all imply a reductionist approach; i.e. treating observational uncertainty as the result of its constituent parts. Alternatively, comparisons between methods provide an estimate of total uncertainty without needing to specify the components. This is especially common when using new technology such as ADCP or LSPIV for discharge (Table 2c) or satellite products for rainfall (Table 1c). Similar comparisons between data sources can be made via water balance analysis to estimate possible rainfall uncertainty (e.g. Heistermann and Kneis, 2011; Kuczera et al., 2010).

The temporal and spatial scales over which uncertainty is calculated are also variable: for example in well-instrumented catchments, instantaneous estimates of uncertainty may be required. However, for a substantial proportion of the globe, access is difficult and irregular, or measurements are only possible by remote sensing techniques. Here, uncertainty estimates integrated over longer time or space scales may be acceptable, e.g. uncertainty in mean annual discharge (Clarke, 1999; Clarke et al., 2000). Even so, without the ability to repeat observations or compare measurement techniques, the challenge of uncertainty estimation and quality control is substantial; for example refer to the discussion by Widen-Nilsson et al. (2007) on data problems when attempting to apply a global water-balance model. Given the recent increase and interest in large-scale hydrological applications, improvements in the knowledge of uncertainty will be of great benefit to improve inter-comparison of data from a wealth of international sources.

3 Rainfall Uncertainty

PERSPECTIVES Uncertainty in rainfall data measured by a gauge network is dominated by (1) point measurement error and (2) spatial variability when interpolating and extrapolating. Lesser uncertainty sources may also be present, and high-resolution temporal variability may exist which is not captured by typical raingauge readings (e.g. every 15 minutes). The literature relating to rainfall uncertainty is well developed, particularly in the case of point measurement error, and includes extensive guidance on quality control of data (e.g. WMO, 1994; WMO, 2008b). Optimal gauge network design to minimise areal mean errors has also been well studied. Increasingly, direct areal measurements are available through radar, leading to different uncertainties (see review by Villarini and Krajewski, 2010). Rainfall measurement uncertainties propagate to cause uncertainty in derived rainfall statistics such as depth-duration-frequency curves (e.g. Molini et al., 2005a; 2005b).

Significant scale differences exist between catchment-based studies and applications for which precipitation estimates are required over scales of thousands of square kilometres, with often sparse and unreliable point estimates available (Steiner, 1996) or relying on alternative data sources such as satellites (Astin, 1997; Stephens and Kummerow, 2007). Global precipitation datasets are in demand for earth system studies but may show significant discrepancies (Fekete et al., 2004). The problem is particularly severe in mountainous areas where steep precipitation gradients exist (Legates and Willmott, 1990) and a large percentage of precipitation falls as snow. The larger uncertainty in mountain precipitation is in conflict with the hydrological significance of mountains as 'water towers' providing freshwater to downstream populations (Viviroli et al., 2003). Meteorological processes are also relevant: uncertainty is large under convective storm cells which can produce large rainfall volumes with very limited spatial and temporal extent. Such events are significant for hazards such as flash flooding, but may be missed altogether from point gauge records and cause difficulties for gauge-radar comparisons (e.g. Rossa et al., 2010; Vasiloff et al., 2009).

Effects of rainfall uncertainty on model calibration and simulation have been studied by means of stochastic simulation (e.g. Andreassian et al., 2001; Bárdossy and Das, 2008; Pappenberger et al., 2005; Younger et al., 2009), or by incorporating rainfall error models into the total error models of uncertainty frameworks (Goetzinger and Bárdossy, 2008; and see Section 6.3 for further references). Conversely, rainfall-runoff models may be used in conjunction with discharge measurements to benchmark the quality of rainfall estimates (Heistermann and Kneis, 2011). Theoretical analysis of rainfall uncertainty requires a prescribed error model, validation of which is relatively rare (Kavetski et al., 2003; 2006b; McMillan et al., 2011b; Moulin et al., 2009; Villarini and Krajewski, 2008; Willems, 2001). Where error model parameters are estimated simultaneously with model parameters, compensational effects between error sources are increased: to avoid this, independent inference of error structures is needed.

In all, recent studies have shown that errors in runoff predictions are often dominated by rainfall bias (Wagener et al., 2007; Yatheendradas et al., 2008) and where uncertainties in rainfall data are recognised, accuracy in hydrological model predictions can be improved (Reichert and Mieleitner, 2009).

TYPICAL RESULTS Refer to Table 1a-1c for quantitative examples.

Point measurement errors

Point measurement errors are dependent on the type of raingauge used, for example: storage, weighing, tipping bucket, drop counting, impact sensor, optical; of these, the tipping bucket gauge is the most widely used. There is a great deal of literature on the causes, effects and corrective procedures for point measurement error, starting from the observations of Heberden (1769). Since then, Sevruk (e.g. 1982; 1996) and Yang (e.g. Yang et al., 1998), amongst many others, have written extensively on systematic errors in precipitation measurements according to gauge type, and correction of the same, although Sieck et al. (2007) showed that common correction techniques fail to account fully for wind-related undercatch.

The figures in Table 1a demonstrate a consensus that systematic undercatch errors are typically in the range 5-16% (Figure 1a). Perhaps the key point, as lamented by Sevruk (1987), is that corrections are rarely made despite this consensus, highlighting the challenge ahead in more poorly understood cases. There may be potential to learn from other disciplines where treatment of bias is better developed (e.g. Magnusson and Ellison, 2008). More recent publications (Ciach, 2003; Krajewski et al., 2003) have analysed random rather than systematic error components, e.g. using data from clusters of raingauges, leading to estimates of mean uncertainty typically around 5% (Figure 1b), and hence of lower magnitude than systematic errors at this point scale (Figure 1a).

Sampling errors/interpolation

For applications using raingauge data in hydrological modelling, estimates of precipitation over entire catchments are required, necessitating interpolation and/or extrapolation from a limited number of point-scale gauges. This process introduces uncertainty into the areal mean rainfall depth calculated at the catchment, sub-catchment or model grid scale. Even in the case of dense gauge networks, variability at small scales of (10^2 – 10^3 m) has been shown to be significant (Goodrich et al., 1995; Wood et al., 2000). Although the metrics used are not always directly comparable, these studies show a rapid increase in rainfall uncertainty with scale, from 4-14 % variation at 10^2 m scale to standard errors of 33-45 % at the 10^3 m scale and 65 % at the 10^4 m scale (Figure 1b). Such figures quickly outstrip the point measurement errors noted in the previous paragraph at all but the smallest scales.

Network design is an important factor which controls interpolation uncertainty (Bras and Rodriguez-Iturbe, 1976; Morrissey et al., 1995; Rodríguez-Iturbe and Mejía, 1974). Estimation of areal mean errors via sub-sampling of a network was used by Horton (1923) and many other since: lists of examples are available in Melching (1995, p79) and Moulin et al. (2009, p100). Geostatistical methods such as kriging also provide estimates of error in mean areal rainfall (Moulin et al., 2009; Storm et al., 1989). More complex approaches to modelling spatial rainfall fields and the associated errors include conditional simulation, i.e. generating ensemble rainfall fields conditioned on the mean and error of spatial rainfall interpolations (Clark and Slater, 2006); note that conditional simulation has also been used for radar and satellite rainfall fields (e.g. Hossain and Anagnostou, 2006; Villarini et al., 2009).

QUANTITATIVE PRECIPITATION ESTIMATION (QPE) FROM RADAR

Weather radar coverage continues to increase in populated areas, and offers the opportunity to sample rainfall rates at high temporal and spatial resolutions, avoiding the interpolation errors discussed above. Instead, radar brings a different set of uncertainties, principally regarding the Z-R (reflectivity – rainrate) relationship, the difficulties in distinguishing solid precipitation, and the effects of terrain blocking. These uncertainties, recently reviewed by Villarini and Krajewski (2010), are currently such that radar is considered most useful as an addition to a gauge network rather than an alternative. Error quantification currently available is usually given as standard deviation of the error as a proportion of rain rate, with typical figures of 0.3 – 0.5 (Table 1c). Errors are shown to be highly dependent on the scale over which areal mean rainfall is required (i.e. larger averaging area reduces the uncertainty); scales of ~1 km can produce standard deviations of 100 % of rain rate (Seo and Krajewski, 2010). Perhaps due to the more obvious uncertainties associated with precipitation estimates from radar, the hydro-meteorology community has quickly established programmes to address the issue (Rossa et al., 2010; Zappa et al., 2010). Progress is being made towards definition of an error model for precipitation estimates from radar, including error model dependency on rain rate, accumulation time, distance from radar and rainfall type (e.g. Ciach et al., 2007; Gebremichael et al., 2011; Kirstetter et al., 2010; Seo and Krajewski, 2010) and error spatial covariance (Berenguer and Zawadzki, 2008; Mandapaka et al., 2009).

4 River Discharge Uncertainty

PERSPECTIVES Recognition of river discharge uncertainty was founded in the hydrometric sciences, where quantification of uncertainty ensured the quality of work. ISO standards provide guidelines for acceptable errors in individual hydrometric measurements (e.g. in velocity-area methods; ISO, 1997) and enable the estimation of combined discharge uncertainty (e.g. Herschy, 1978). A recent, thoughtful review by Hamilton and Moore (2012) describes how uncertainty in published streamflow records changes due to technical and methodological advances, and examines the implications for uncertainty reporting. When hydrologists make use of hydrometric data for modelling applications, discharge uncertainty is always present, and becomes influential during calibration and validation. Recent studies comprise several in which error models for discharge have been *justified* and included (Huard and Mailhot, 2008; Krueger et al., 2010a; Liu et al., 2009; McMillan et al., 2010; Pappenberger et al., 2006; Westerberg et al., 2011). Discharge uncertainty has also important implications for interpreting hydrometric data to describe and understand catchment dynamics, and this is discussed more fully in Section 6.1.

TYPICAL RESULTS The characteristics of uncertainty in river discharge are entirely dependent on the technique used to measure and/or compute the discharge. The technique chosen naturally depends on a range of factors including the flow volume, channel characteristics, ease of access to the channel and application-specific considerations such as funding availability and length and accuracy of the measurement series required. The most common method for discharge calculation remains the use of a rating curve to convert measured stage to a discharge value (refer to Schmidt (2002) for a historical background). The resulting discharge uncertainty can be decomposed into several distinct error sources. A significant body of work is available to quantify the individual and combined uncertainties and provides an excellent reference against which to frame any application. The components are introduced next; refer to Table a-2c for quantitative results.

STAGE/DISCHARGE MEASUREMENTS (AREA, VELOCITY, STAGE)

Each point on a rating curve is composed of a stage measurement at the recording gauge, and a discharge measurement at a nearby cross-section. The uncertainty in the stage measurement is generally considered to be small (e.g. Dymond and Christian, 1982). Van der Made (1982) and Petersen-Øverleir and Reitan (2005) provide a summary of the components of this uncertainty, and estimates relating to specific measurement techniques (e.g. stilling well floats) are also available. Instrument precision values are usually given as a range in mm, and rarely exceed ±10 mm (Figure 2a). The discharge measurement is naturally more uncertain and can be prone to errors associated with both instrumentation and quality control (especially in more remote sites) potentially leading to outliers (Sefe, 1996). Historically, the most common procedure involves calculation of the mean velocity at a cross-section using discrete current meter measurements across the width and depth profile of the river. Pelletier (1988) provides a comprehensive literature review of the uncertainties arising from this calculation, depending on factors such as velocity, time of exposure and number of verticals used. Total uncertainties at the 95% confidence level were found to lie in the range 4-17% for 5-35 verticals. More recently, Herschy (2002) uses figures from the ISO standard 748 to reach an estimate of the combined uncertainty at 6% (Figure 2b).

INTERPOLATION/EXTRAPOLATION OF RATING CURVE

Uncertainty as to the true form of the stage-discharge relationship leading to its approximation by fitting a rating curve using interpolation or a functional type (e.g. power law) is a major source of uncertainty in discharge estimation. Errors increase if the rating curve is extrapolated beyond the observed stage-discharge measurements (Kuczera, 1996; Mosley and McKerchar, 1993). Clearly, the approach taken to extrapolate should consider the cross-section stability (i.e. fixed structure vs. rated section), geometry and control for high flows. In controlled cases extrapolation may be simplified, e.g. by the use of Manning's equation where Manning's roughness is not expected to vary significantly. Venetis (1970) was among the first to recommend that rating curve uncertainties should be treated within a statistical framework. The classical statistical approach has been rigorously updated by Moyeed and Clarke (2005), Petersen-Øverleir and Reitan (2009) and Reitan and Petersen-Øverleir (2006; 2008; 2009) using a variety of techniques including multi-segment fitting and Bayesian estimation. Fuzzy methods have been favoured by other authors using, for example, 'limits of acceptability' approaches (Liu et al., 2009), envelope curves (Krueger et al., 2010a; Pappenberger et al., 2006) and fuzzy set theory (Shrestha et al., 2007; Shrestha and Simonovic, 2010a; b). Empirical methods such as multipliers for the rating curve have been used as well (Aronica et al., 2006). The rating curve approach has also been extended to account for uncertainties due to unsteady flow, e.g. by including longitudinal variation in water surface slope using simultaneous stage measurements at two adjacent cross sections (Dottori et al., 2009; Dottori and Todini, 2010; Koussis, 2010; Leonard et al., 2000; Schmidt and Yen, 2008). In large river systems especially, extrapolations may be needed in both directions (high and low flows) from gauged relationships for effective resource management (Sefe, 1996).

Channel Cross-Section Change

Changes in morphology and channel cross-section introduce further uncertainties into discharge measurements. Multiple causes exist, including seasonal ice cover at the gauging site (e.g. Shiklomanov et al., 2006; and a review by Pelletier (1990)), or seasonal vegetation growth. Jalbert et al. (2011) used variographic analysis to model the resulting increase in discharge uncertainty with time. Rating changes due to channel morphology are particularly pronounced in alluvial rivers (Burkham and Dawdy, 1970). Under unsteady flow, a coupled relationship between the evolution of river bed forms and the stage-discharge relationship can occur (Shimizu et al., 2009). In many of the world's large and dynamic rivers, complex changes in river morphology are common (Ashworth et al., 2000; Goswami et al., 1999; Sarma, 2005). For example in a study of the large Brahmaputra river at Bahadurabad in Bangladesh it was estimated that the combination of changing bed forms and inaccurate measurements of velocities and current meter depths could cause up to 20 % uncertainty in discharge measurements (Mirza, 2003). Reitan and Petersen-Øverleir (2011) modelled unstable rating curves by evolving the parameters of the standard power-law stochastically over time as part of a hierarchical Bayesian error model.

Combined

In many cases the total discharge uncertainty will relate to a combination of all the sources cited above. Some authors have undertaken studies which aim to quantify this combined uncertainty. Di Baldassarre and Montanari (2009) considered uncertainties resulting from point measurement error, hysteresis (see also Perumal et al., 2004), roughness change due to vegetation growth, and extrapolation; all calculated as percentage errors and then combined for total discharge uncertainty. McMillan et al. (2010) combined uncertainties including point measurement, rating curve form, and cross-section change in a gravel-bed river to derive a complete probability distribution function of discharge for any measured stage value. Westerberg et al. (2011) also considered the instability of rating curves, in conjunction with gauging measurement error and rating curve extrapolation. Authors usually cite confidence bounds for the relative discharge error, making comparison between sites possible, with typical values being ±50-100% for low flows, ±10-20% for medium or high (in-bank) flows, and a single estimate of ±40% for out of bank flows (Figure 2b).

REPRESENTATIONAL UNCERTAINTIES

Insufficient temporal sampling of stage in calculation of mean discharge, e.g. daily, is an additional source of uncertainty, especially where peak flows have short duration (Petersen-Øverleir et al., 2009; Westerberg et al., 2011). An extreme example of restricted temporal coverage occurs where subjective interpolation methods are used to estimate winter discharge under ice, with results highly dependent on incorporating process knowledge such as flow reduction during ice formation periods (Hamilton and Moore, 2012). In addition to measurement error, discharge data may have a systematic bias due to unknown water losses circumventing the gauge. This bias is likely to have greater significance at low flows. Further representational uncertainties may be induced by uncertainty in and dynamics of the catchment area of a gauge (Krueger et al., 2010b) as well as unknown gain or loss fluxes when calculating water balances for modelling or field studies (Genereux et al., 2002; Graham et al., 2010). Importantly, there are almost no studies or information about the potential for such errors in different catchments.

ALTERNATIVE TECHNIQUES

New instruments and techniques continue to be developed to measure discharge more directly rather than using a rating curve, for example velocity measurement using acoustic Doppler velocimetry (ADV), acoustic Doppler current profiling (ADCP) or Large Scale Particle Image Velocimetry (LSPIV). These techniques have their own individual uncertainty characteristics, and methods to quantify the uncertainty are still under development in many cases. The World Meteorological Organization currently leads a research programme to design and collate appropriate uncertainty estimation methods (WMO, 2008a; 2010). Some studies have derived uncertainty estimates by comparison with impellor or other standard measurements (which are themselves subject to uncertainty); e.g. McIntyre and Marshall (2008) for ADV, Oberg and Mueller (2007) for ADCP, Costa et al. (2006) for radar. The corresponding uncertainties found, given as relative error, were ±20 % (range; ADV), ±3-7 % (standard deviation; ADCP) and larger values of 2-47 % (range; LSPIV) (Figure 2b). A current review of PIV techniques and uncertainties is provided by Muste et al. (2008).

In many parts of the world, rivers are not accessible for direct gauging even where estimates of discharge are required, and remote sensing techniques are used. They may include measurement of channel width, water elevation and velocity from satellites (Bjerklie et al., 2003; Birkinshaw, 2010; Negrel et al., 2011) or aerial photography (Bjerklie, 2007; Bjerklie et al., 2005). These methods have great potential for use in global discharge estimates but are naturally highly uncertain and so it will be important to understand the potential for accompanying errors.

5 Water Quality Uncertainty

PERSPECTIVES Water quality data typically subsumes concentrations of solutes and solids such as industrial and agricultural pollutants and derived quantities thereof (e.g. loads). In order to make the scope of the paper manageable, we do not deal with microbial concentrations and biological indicators here, but these are equally affected by uncertainties (Rode and Suhr, 2007; Schmidt and Emelko, 2010). We further limit our focus to suspended solids (SS), phosphorus (P) and nitrogen (N), which are the most documented water quality determinants.

The explicit consideration of data uncertainty in water quality research is a more recent endeavour than the consideration of rainfall and discharge uncertainty discussed so far, although early quantitative research on sampling uncertainty appeared in the fluvial sediment transport literature (Horowitz et al., 1990; Walling and Teed, 1971). Ward (1984) added an analytical perspective by describing vertical sediment sampling uncertainty in a cross-section using hydraulic theory, albeit assuming ideal conditions. In P research an early discussion of data uncertainty appeared in Reckhow and Chapra (1979), while Haan (1995) was among the first to call explicitly for quantitative uncertainty assessment in both collection and reporting of data. Two comprehensive quantitative reviews of uncertainties in selected water quality data have appeared since (Harmel et al., 2006; Rode and Suhr, 2007), with the Harmel et al. data and error propagation method available as a software tool (Harmel et al., 2009). However, including uncertainty analysis in field and modelling studies is still far from common practice. This deficit compromises the determination of margins of safety for water quality protection (Harmel et al., 2006), the identification of trends, and the proper driving and evaluation of models (Johnes, 2007; Krueger et al., 2007; Radcliffe et al., 2009; Rode and Suhr, 2007).

Many models used in water quality research are statistical models (e.g. Plate, 1995) and thus rely explicitly on the quality of the data. Export coefficient type models are equally common, and require calibration against loads which are derived data products particularly affected by the representational uncertainties discussed below. Mechanistic models, too, rely on the quality of the calibration data. However, only a few studies have addressed the issue of data uncertainty in modelling explicitly. McIntyre et al. (2002) and McIntyre and Wheater (2004) explored the effect of water quality data uncertainty and sampling frequency on calibration and prediction of mechanistic models in numerical experiments assuming idealised synthetic errors. McIntyre and Wheater (2004) in particular showed differences in calibration and predictive performance of a river transport model for different P load data scenarios and the limited value of routine low frequency P sampling for driving and calibrating the model. While both studies relied on idealised errors, estimating actual error structures from typically available data has rarely been possible in water quality modelling to date (McIntyre et al., 2003), hence subjective assumptions have been inevitable. McIntyre et al. approximated the uncertainty in daily averages of a number of water quality variables as a uniform distribution between the minimum and maximum of three daily samples. These data were then used as part of an uncertainty analysis of a river water quality model. Beven et al. (2006) also defined P concentration data uncertainty subjectively and used these as limits of acceptability for rejecting competing parameterisations of a P leaching model within the extended Generalised Likelihood Uncertainty Estimation (GLUE) methodology. Krueger et al. (*in press*) followed the same methodology in a comparison of empirical SS and P transport models, but were able to relate their data uncertainty estimates to the number of sub-samples per timestep for at least a few timesteps where these data were available. Harmel and Smith (2007) incorporated data uncertainty estimated based on the quantitative review of Harmel et al. (2006, summarised in Table 3b) into performance metrics of a P transfer model.

TYPICAL RESULTS Concentrations are the basis of most derived data products, and are typically determined by analysing water samples (in the lab or in the field) which are taken manually or automatically. Some determinants can now be measured by in-situ probes either directly or indirectly (e.g. SS via turbidity). The levels of uncertainty are discussed next and Tables 3a-3c list typical quantitative results for SS, P and N.

MEASUREMENT UNCERTAINTIES

Data based on sampling are affected by errors in capturing a volume of flow without altering its composition of dissolved and suspended substances (Wass and Leeks, 1999). The errors associated with automatic sampling systems may be greater in this respect compared to manual samples, as these rely on sufficient suction and prior flushing and are confounded by clogging, biofouling and, for particles and particulate-bound substances, preferential sampling effects (Evans et al., 1997; Jordan et al., 2005). For SS concentrations, for example, sampler effects may introduce a relative error as large as 36 % (Figure 3a). For storm loads, relative sampling error can be 14-33 % for SS and 0-17 % for total P, but around zero for total N which is mostly un-affected by preferential sampling effects (Figure 3). For manual sampling, relative error contributions to SS storm load of 15-50 % have been reported (Figure 3a).

If samples are not analysed directly on-site (via bank-side analysers; Jordan et al., 2005) then biogeochemical transformation processes may further alter their composition during transport and storage (Kotlash and Chessman, 1998; Robards et al., 1994; Worsfold et al., 2005). These effects are particularly prominent when sampling P, and can introduce a relative error of 64-92 % into total P storm loads (Figure 3b). Additional errors may be introduced by sub-sampling for laboratory analysis (Donohue and Irvine, 2008) and other sample preparation steps (e.g. Magnusson et al., 2004). Sub-sampling may introduce a relative error of 6-8 % into total P and of 10-11 % into total N concentration measurements (Figure 3b, c).

Finally, both water samples and in-situ probes are affected by the precision of the analytical instruments used (Meyer, 2007), which is generally quoted by manufacturers as below 5 % (see Figure 3b, c for total P and total N concentrations). Individual error components combine to produce analytical uncertainty typically in the order of 5 % (see Figure 3b, c for total P and total N concentrations).

REPRESENTATIONAL UNCERTAINTIES

Measurement uncertainties describe the uncertainty in a determinant concentration at one point in a stream, river or lake. If this point concentration is being related to the larger spatial and temporal scales which are typically of interest, e.g. the determinant load of a stream over the scales of events up to multiple years, then representational uncertainties may be dominant, although in some cases lab analysis can contribute most to the overall uncertainty (Harmel et al., 2006), e.g. up to 9.8 % in SS storm loads (Figure 3a). Representational uncertainties are induced by determinant concentrations varying within a flow cross-section (Horowitz et al., 1990; Lovell et al., 2001; Rode and Suhr, 2007; Wass and Leeks, 1999) and across a lake – scales that are misrepresented to some extent by point measurements (in-situ or via sampling). For SS concentrations, for example, variations in the river cross-section may introduce a relative sampling error as large as 26 %; for instantaneous loads the relative error may be up to 14 % (Figure 3a). In addition, determinant concentrations vary over short time periods (Horowitz et al., 1990; Rode and Suhr, 2007), and the measurement frequency determines the accuracy and precision with which certain temporal dynamics can be resolved (Jordan et al., 2005; Lazzarotto et al., 2005), including the effectiveness of pollution mitigation measures (Jordan and Cassidy, 2011). For example, the uncertainty of SS and total P hourly flow-weighted mean concentration as a function of temporal sampling resolution can range from 10 to 50 % (Figure 3a, b). The variability of particulate substances is generally much larger than that of dissolved ones (Lovell et al., 2001; Rode and Suhr, 2007).

Unrepresentative temporal sampling affects annual frequency distributions of determinants, particularly the loss of right skewness and tails (Johnes, 2007), and associated percentiles for comparison to water quality targets (van Buren et al., 1997). It also affects determinant loads; and the accuracy and precision of various temporal sampling and load estimation strategies have long been researched in the fluvial sediment transport literature and recently in water quality research more widely. The early studies used turbidity-generated high-resolution concentration time series paired with flow time series to construct benchmark loads against which low-resolution sampling and estimation strategies were compared by sub-sampling the time series. Walling and Webb (1981) and Phillips et al. (1999) provide comprehensive summaries. Here we focus on studies that used actual high-resolution concentration data for benchmarking. Temporal resolution is obviously relative to the concentration dynamics of interest. For large rivers daily sampling may be classed as high-resolution (e.g. Al-Ansari et al., 1988; Asselman, 2000; Dolan et al., 1981; Horowitz, 2003), whereas for small rivers and streams high-resolution means sub-daily (e.g. Bowes et al., 2009; Kronvang and Bruhn, 1996; Salles et al., 2008; Stevens and Smith, 1978) or even sub-hourly sampling (e.g. Jordan and Cassidy, 2011).

Inadequate temporal sampling can result in a relative bias of up to 65 % in storm loads (Figure 3). For annual SS loads a bias of up to 30 % has been reported (Figure 3a), for annual total P loads up to 150 % (Figure 3b). Moatar et al. (2006) correlated SS load uncertainty positively with the importance of extreme events for mass transfers, which in turn decreased with increasing basin size. Johnes (2007) correlated total P load uncertainty negatively with base flow index (i.e. uncertainty increased with the importance of extreme events as with SS) and positively with population density (as an index for point sources). Load estimation methods diverge too (Figure 3), yielding biases of up to 52 % in SS storm loads, up to 38 % in total P storm loads, and up to 22 % in total N storm loads, for example. Component uncertainties can compensate each other and generally average out over longer aggregation times, but combined uncertainties are still highly variable (Figure 3): 0-33 % for SS concentrations, 15-35 % for SS storm loads, 16-104 % for total P concentrations, 17-105 % for total P storm loads, 0-10 % for daily total P loads, 14-104 % for total N concentrations, 15-105 % for total N storm loads.

Proxy measurements

For indirect methods, proxy measurements have to be related to the quantity of interest by calibration or a model. This introduces uncertainties associated with this relationship on top of the measurement and representational uncertainties (Foster et al., 1992; Gippel, 1995; Teixeira and Caliari, 2005; Wass and Leeks, 1999; Wass et al., 1997; Eder et al., 2010, for the case of turbidity-suspended solids relationships).

6 Data Uncertainty: Implications

The process of identification, summary and comparison of uncertainties in rainfall, river discharge and water quality variables that we have followed above, has left no doubt that data uncertainties are widespread and of significant magnitude. Therefore it is important to consider how those uncertainties impact on the interpretation of the data in order to draw scientific conclusions. In this section we discuss the impacts on three areas: interpretation of catchment dynamics, model regionalisation and model evaluation.

6.1 Interpretation of Catchment Dynamics

Improvement in the understanding and characterisation of hydrological systems is at the heart of all catchment monitoring programmes. Some programmes are purely field-based, but increasingly modelling is used alongside field campaigns to synthesise and develop new insights into catchment responses (Dunn et al., 2008; Tetzlaff et al., 2008), and inform future data collection (Dunn et al., 2007; McGuire et al., 2007). Uncertainty in field data, therefore, clouds the potential to interpret catchment dynamics in two ways: through direct contamination of observed responses, and indirectly through biasing model predictions and hence compromising the iterative modelling-measuring cycle of improved understanding.

In recent years, developments in measurement technology have allowed an increasing number of alternative data sources to augment rainfall and flow time series in the characterisation of catchment behaviour. Soil moisture content and water table level are now routinely measured, and the latest advances include the use of tracers such as stable isotopes (Birks and Gibson, 2009; Soulsby et al., 2000), chloride (Page et al., 2007), Gran alkalinity (Birkel et al., 2010) and diatoms (Pfister et al., 2009). These new data sources not only bring their own individual measurement uncertainty characteristics; but also typically rely on co-measurement of rainfall and flow data (e.g. flow-concentration curves) for their interpretation, which are affected by the many sources of uncertainty raised in this paper. When multiple data sources are used, data errors may cause incompatibility of process interpretations at the scale of interest (Lischeid, 2008). Soft data (Seibert and McDonnell, 2002) adds additional challenges to uncertainty quantification, although after consideration of the data uncertainty identified in this review the distinction between hard and soft data may become less clear.

The use of models in combination with field data to identify catchment dynamics is particularly susceptible to data uncertainty in regard to fast response processes where errors may be increased due to high flows and flashy behaviour. Examples include identification of variable source areas (Beven and Freer, 2001; Dunne and Black, 1970) or model confirmation of groundwater ridging behaviour (Cloke et al., 2006). Proper representation of processes such as infiltration excess runoff or activation of transient stream channels is dependent on the accuracy of high resolution rainfall data under high intensity conditions, when input uncertainties are likely to peak. When reliable high resolution measurements become available, improved process understanding can be gained, as has been shown for hyporheic process dynamics (Malcolm et al., 2006) and estimates of P load, sources and dynamics (Johnes, 2007; Jordan et al., 2005; Jordan and Cassidy, 2011; Lazzarotto et al., 2005; Radcliffe et al., 2009).

Recently, the opportunity to use multi-response field data to build an integrated view of dominant processes in a catchment has also been harnessed to guide model structure (Clark et al., 2011; Fenicia et al., 2008a; McMillan et al., 2011a). It is a priority to include data uncertainty in such analyses as the diagnostics employed have the potential to be altered by data errors. For example in the context of eutrophication studies, Hanafi et al. (2007) demonstrated that the uncertainty propagated from nutrient concentration measurements through to nutrient uptake length and velocity was too large to distinguish between high and low uptake conditions. Similarly in the study by Kennard et al. (2010), large uncertainties in flow metrics clouded ecological impact assessment. The problem is especially severe when diagnostics rely on relationships between two data streams, for example: water balance analysis (Graham et al., 2010), threshold response in the ratio of precipitation to runoff, recession analysis of the relationship between flow and its derivative. We therefore stress the importance of data uncertainty analysis for robust interpretation of catchment dynamics.

6.2 Model Regionalisation

The effect of uncertainties in observed data will propagate from process identification and model structural choice into wider fields such as model regionalisation and predictions in ungauged basins. Bai et al. (2009) noted the role of data uncertainty in addition to parameter uncertainty in top-down watershed model evaluation. Wagener and Wheater (2006) highlighted the path by which model structural uncertainty could lead to non-identifiability of catchment model parameters, and hence uncertainty in the regionalisation method. Attempts have been made to address some of the uncertainty sources, e.g. Yadav et al. (2007) included uncertainty estimation within the regression equations of their regionalisation approach, and McIntyre et al. (2005) used multiple models within the regionalisation to account for structural uncertainty.

However, these methods currently lack a full analysis of observational uncertainties, which are instead implicitly mapped onto parameter uncertainties with the attendant risk of bias in parameter estimates and model forecasts. Regionalisation of hydrological models also implicitly assumes the comparability of measurements between basins; which will not be the case as data uncertainties vary spatially. For example, discharge data may be used to evaluate spatial patterns of model parameters or flow recession characteristics, while their uncertainties depend strongly on the range of channel types and the data collection method, e.g. via natural rated sections or gauging structures.

6.3 Model Evaluation

The potential of incorrect uncertainty assumptions to cause bias in model calibration and predictions (as discussed by Beven et al., 2007; 2011; Ibbitt, 1972; Kavetski et al., 2006a; 2006b; Sorooshian, 1981; Thyer et al., 2009; Troutman, 1982; 1983; Vrugt et al., 2008) has led to calls for more thoughtful approaches to uncertainty estimation. While there is an increased acceptance of data uncertainty in the modelling community, hydrological modellers are often reliant on the analysis and provision of error structures and magnitudes alongside field data sets (Graham et al., 2010).

Model calibration schemes which treat model input and/or output uncertainty explicitly demonstrate an ability to incorporate, and a growing requirement for, sound advice on measurement uncertainty magnitude and form. Such schemes include BATEA (Kavetski et al., 2003; 2006a; 2006b), DREAM (Schoups and Vrugt, 2010; Vrugt et al., 2008), IBUNE (Ajami et al., 2007) and extended GLUE (Beven et al., 2006; Beven, 2006; Krueger et al., 2010a; in press; Liu et al., 2009; Pappenberger et al., 2006; Quinton et al., 2011). Accordingly, an increased number of studies specify at least one measurement error model (Huard and Mailhot, 2006; 2008; Kennedy and O'Hagan, 2001; McMillan et al., 2010; Reichert and Mieleitner, 2009; Vrugt et al., 2003; 2005; Vrugt and Robinson, 2007). Harmel and Smith (2007) modified traditional model performance metrics with data uncertainty information, while Khadam and Kaluarachchi (2004) incorporated qualitative information on the reliability of data. The expected improvement of model performance metrics when data uncertainties are included explicitly is balanced with increased equifinality (Krueger et al., 2009). However, without including data uncertainties the performance metrics may be intrinsically compromised, and unsuitable for their fundamental purpose of making comparisons across dimensions of model structure, parameterisation, time and space.

Estimation of the complexity of dominant processes and hence appropriate models is also subject to interference from data errors; whether this is achieved through analysis of time series (Sivakumar, 2004) or a top-down modelling approach (Fenicia et al., 2008b; Klemeš, 1983). Singh and Woolhiser (1976) found that large input errors could overwhelm the non-linear surface runoff responses of their model and so a simpler linear model became the preferred choice. In the context of model identification, an additional source of uncertainty lies in the information content of observed data: is the series consistent with the overall observed responses, and are the conditions observed sufficient to excite the full range of model responses (Gupta and Sorooshian, 1985; Sorooshian et al., 1983; Young, 2003).

7 Towards a Culture of Working with Data Uncertainty

7.1 Summary of Findings

The information that we have reviewed, contextually and in Tables 1-3, has allowed a comparative study of different observational uncertainty sources. In particular, we have been able to identify dominant uncertainties in raw measurements and in derived quantities relevant to different scales, processes, scientific questions and disciplines.

In the case of rainfall, we saw an overarching narrative in which uncertainty magnitudes and our ability to characterise them were driven by scale. At the point scale, uncertainties were comprised of systematic, usually undercatch, errors of average magnitudes 5-16 % which could in theory be subject to correction (Figure 1a), plus random errors of magnitude around 5 %. However, as the need for estimates of areal mean rainfall was scaled up, interpolation errors came into play which could vary from 4-14 % variation at the 10^2 m scale to standard errors of 33-45 % at the 10^3 m scale and 65 % at the 10^4 m scale. At larger scales of 10^3 to 10^5 m, radar or satellite estimates of rainfall could have uncertainties of 9-150 % of rainfall rate, whose magnitudes were however reduced by averaging over larger areas. Timescales also affect uncertainties which are reduced with longer averaging times, although the gradient may not always be consistent when comparisons are made across different studies. These findings are summarised in Figure 1b.

For discharge uncertainty, scale was less important due the integrated nature of the measurement. Instead, the gauging method used together with the relative flow (low flow up to flood) was key to understanding uncertainty (Figure 2). Individual measurements of discharge have uncertainties in the range 2-19 % using velocity-area methods, with similar ranges for the newer methods of ADV and ADCP, and 2-47 % range for LSPIV measurements. However, once extrapolation must be made outside of the stage range or channel conditions used for gauging, rating curves are typically used for discharge estimation. This method brings much larger uncertainties and is the main contributor to estimates of total discharge uncertainty with confidence bounds of typically ±50-100 % for low flows, ±10-20 % for medium or high (in-bank) flows, and ±40 % for out of bank flows. The last figure is based on only one reference and for more extreme floods larger uncertainties can be expected, though currently not quantified.

Water quality data uncertainty is highly variable as it results from a combination of a larger number of component errors compared to rainfall and discharge, which combine differently for different environments, methods, types of equipment and seasons. Uncertainty also aggregates differently for average concentrations and loads over different timescales due to the fundamental representational limitations imposed by the need for spatial and temporal sampling of constituents, with a tendency for component errors to average out with aggregation time (Figure 3). Combined analytical uncertainty can generally be considered smallest, in the order of 5 %, though with exceptions, e.g. for SS. Sample timing effects are typically greater than effects of the actual sampling method and conduct. For example up to 65 % for storm loads; compared to 14-33 % for SS, 0-17 % for total P and around zero for total N, also highlighting the greater susceptibility of particulate/particulate-bound substances (such as SS and total P) to preferential sampling effects and cross-sectional variation.

7.2 Guidance

Based on the evidence presented in this paper, we venture to make a number of recommendations for research into observational uncertainty. These highlight some important issues that have been raised in relation to existing gaps in our understanding and treatment of observational uncertainty. We advocate a culture of explicitly acknowledging and working with uncertainty, not least to eventually demonstrate appreciable uncertainty reduction.

1. **Community sharing of information.** Combine our knowledge of error characteristics and magnitudes for different data sources relevant to hydrology, while recognising and describing the place-dependency of some error types. Also catalogue the potential for as-yet unmeasured uncertainties in those data sources.

2. **Uncertainty as metadata.** Develop metadata standards that fully characterise data uncertainty. For example, the hydrological data standard proposed by Tarboton et al. (2008) allows the association of a single ‘Value Accuracy’ with each data point, but does not differentiate between different causes or types of uncertainty, such as bias vs. precision. The HarmoniRiB project designed a database for river basins in the context of delivering the EU Water Framework Directive with scope to associate a comprehensive probability model with each uncertain data item (Refsgaard et al., 2005). An additional semi-qualitative description of the ‘pedigree’ of data, describing the limits of the state-of-the-art in producing these data, could be based on the NUSAP notational system (Constanza et al., 1992).

3. **Characterisation of uncertainty.** Improve knowledge of error distributions, often lacking in existing data. This applies to both raw measurements and to integrated or derived quantities (e.g. point vs. areal mean rainfall), and should include explicit recognition of commensurability errors when comparing models to data. Uncertainty information may be needed at a sub-data series level (e.g. different discharge ranges may relate to different measurement techniques: wading, cable, ADCP).

4. **Training in observational uncertainty.** Include exposure to concepts of data uncertainty within hydrological sciences training programmes, helping to develop good practice in working with and reporting data uncertainty.

5. **Learning through dialogue.** Improve the dialogue with the statistical community when developing guidance about appropriate uncertainty analysis techniques. Improve the dialogue between experimentalists and modellers to encourage consideration of the effect of experimental design on uncertainty in the required derived quantities. Improve the dialogue with the water management community to understand and foster user demand for uncertainty information alongside hydrological data.

6. **Learning through experimentation.** Greater emphasis on observational uncertainty experiments in project proposals will enable us to assess how different uncertainty characterisation methods affect model predictions, uncertainty bounds and performance. Such experiments will be needed to understand the value of new and diverse hydrological data sources, their specific information content, and the level of complexity in our methods needed to provide error estimates.

7. **Design of diagnostics.** Where observations are used as diagnostics, e.g. for model evaluation or in decision frameworks, those diagnostics should be robust to some level of observational uncertainty. This might imply greater use of integrated diagnostics (e.g. the annual rainfall-runoff ratio) over timestep-based measures such as those based on the sum of squared errors, which are sensitive to time-varying random errors. Alternatively, comparative or change-based diagnostics may allow defensible decision-making under uncertainty.

More broadly, this review has highlighted how knowledge of data uncertainties is important in the calculation of hydrological indices, summary statistics, process interpretation and model predictive capability. Quantitative uncertainty estimates are needed for communication of data uncertainty across disciplinary boundaries, to data users, policy makers and to the general public. Uncertainty quantification is prerequisite to understanding how much information in needed to characterise system behaviour before catchment response and hydrological change signals can be separated from natural variability and measurement uncertainty (Kennard et al., 2010; Burt et al., 2010). Good quality data is essential: high-resolution monitoring is valuable to quantify representational uncertainties that are critical to understand before we can regionalise process knowledge and models. Similarly, good quality, long term data sets facilitate emerging tools which identify previously uncharacterised data errors by analysis of 'unusual events' (e.g. regression tree analysis, Ali et al., 2010; data depth functions, Singh and Bardossy, 2012). As described by Hamilton and Moore (2012), routine uncertainty reporting enables the value of high-quality data and post-processing to be recognised by data users, and encourages best practice by data providers to reduce data uncertainties.

However, it is unreasonable to expect every study to allow the funds and time required for in-depth study of individual observational data types, especially at scales relevant for policy and management which are notoriously difficult to cover. Therefore a key requirement is for a wider analysis and synthesis of data errors to provide *a priori* guidance. We have started the process in this review, but recognise that currently some of the conclusions we can draw about the distribution of error characteristics are weak; for example we may only be able to provide plausible upper bounds for an uncertainty type, or a summary of uncertainties encountered in previous experiments. In particular, the number of 'replicate' studies we were able to summarise here was very low.

To address these limitations it is essential to encourage good practice in reporting data uncertainty: we found that it was not always possible to extract the exact uncertainty metrics used in published studies. The ability to share, synthesise and re-use information will be greatly enhanced if published uncertainty estimates are more precise. For example, when uncertainty ranges, standard deviations or confidence intervals are reported, these should be accompanied by a description of the method used for calculation including the underlying distribution assumption where appropriate. Standardised reporting would start the process of attribution of data error characteristics and dominant uncertainties. Factors which control the uncertainties will depend on data type (refer to Figures 1-3) and will in some cases be place-dependent. Understanding these factors will help to constrain errors where there is no local information.

This type of synthesis will become more important as we increasingly rely on large shared data information systems (e.g. the CUAHSI Hydrologic Information System; Tarboton et al., 2010), and modellers may retreat further from field data collection. According to one reviewer, neither of the two hydrological observatories HOBE (Denmark) and TERENO (Germany), which aim to produce high quality data for scientific purposes, systematically store data with uncertainty information. In the UK, due to our involvement in the Demonstration Test Catchments programme, some uncertainty data has been recorded for future inclusion in the project database, but again is not a funding priority and is made more difficult by the lack of a common standard for uncertainty in hydrological data. This point can be further emphasised by the increasing interest in large scale hydrology and model comparison with global discharge products such as those obtained from the Global Runoff Data Centre (http://grdc.bafg.de) and Composite Global Runoff Fields (e.g. Fekete et al., 2002). Albeit such observations and data products are very welcome to quantify hydrological simulations over diverse environments, the quality and validity of such information (including metadata such as station co-ordinates) varies significantly and is relatively unreported and unknown (Pappenberger et al., 2010). We therefore risk becoming disconnected from the interpretation of data quality unless it is properly embedded in any metadata information abstracted from such data information systems, or methods are available to estimate uncertainties where they are not available directly.

A community discussion on data uncertainty has the potential to lead to clearer mechanisms for sharing knowledge and impacts, and build consensus for the reporting and propagation of data uncertainties. More effective sharing of existing data and collaborative reflection on uncertainty estimation methods will reduce the danger of propagating artificial levels of confidence (Brown, 2004). We have started this process by posting the tables of this paper on the Experimental Hydrology Wiki[1] and encourage readers to contribute their own findings. This resource and tools like those of Brown and Heuvelink (2007) and Harmel et al. (2009) provide a starting point for scoping data uncertainties, and the reference to common sources and tools will lend transparency and repeatability to uncertainty quantification. However, it is important to stress that such *a priori* information should be augmented with independent quantitative evidence as this becomes available on site. Such a learning process sits comfortably with the Bayesian mode of inference that has been gaining acceptance in hydrology. We explicitly discourage a 'job done' mentality in cases where reference to existing sources and tools on data uncertainty is made.

[1] www.experimental-hydrology.net

ACKNOWLEDGEMENTS

During writing of this paper, HM was funded by the NZ Ministry of Science and Innovation (grant no. C01X0812). TK was funded by the UK Research Councils' Rural Economy and Land Use (Relu) Programme (grant no. RES 229-25-0009-A) and a UK Natural Environment Research Council Knowledge Exchange Fellowship (grant no. NE/J500513/1). JF was funded by the UK Natural Environment Research Council (grant no. NE/1002200/1) and the UK Department for Environment, Food and Rural Affairs (grant no. WQ0211). We thank Ida Westerberg, Paul Smith and an anonymous reviewer for their thorough and constructive reviews of this manuscript.

REFERENCES

Ackers, P., 1978. Weirs and flumes for flow measurement. Wiley, Chichester.
Ajami, N.K., Duan, Q.Y., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Research, 43(1): W01403.
Al-Ansari, N.A., Asaad, N.M., Walling, D.E., Hussan, S.A., 1988. The suspended sediment discharge of the River Euphrates at Haditha, Iraq: An assessment of the potential for establishing sediment rating curves. Geografiska Annaler, Series A, Physical Geography, 70(3): 203-213.
Ali, G. A., Roy, A.G., Turmel, M.C., Courchesne, F., 2010. Multivariate analysis as a tool to infer hydrologic response types and controlling variables in a humid temperate catchment. Hydrological Processes 24(20): 2912-2923.
Andreassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. Journal of Hydrology, 250(1-4): 206-223.
Andréassian, V., Perrin, C., Parent, E., Bárdossy, A., 2010. The Court of Miracles of Hydrology: can failure stories contribute to hydrological science? Hydrological Sciences Journal, 55:6, 849-856
Aronica, G., Candela, A., Viola, F., Cannarozzo, M., 2006. Influence of rating curve uncertainty on daily rainfall-runoff model predictions. In: Sivapalan, M., Wagener, T., Uhlenbrook, S., Zehe, E., Lakshmi, V., Liang, X., Tachikawa, Y., Kumar, P. (Eds.), Predictions in Ungauged Basins: Promise and Progress. IAHS Publications 303, pp. 116-124.
Ashworth, P.J., Best, J.L., Roden, J.E., Bristow, C.S., Klaassen, G.J., 2000. Morphological evolution and dynamics of a large, sand braid-bar, Jamuna River, Bangladesh. Sedimentology, 47(3): 533-555.
Asselman, N. E. M., 2000. Fitting and interpretation of sediment rating curves. Journal of Hydrology, 234(3-4): 228-248.
Astin, I., 1997. A survey of studies into errors in large scale space-time averages of rainfall, cloud cover, sea surface processes and the earth's radiation budget as derived from low earth orbit satellite instruments because of their incomplete temporal and spatial coverage. Surveys in Geophysics, 18(4): 385-403.
Bai, Y., Wagener, T., Reed, P., 2009. A top-down framework for watershed model evaluation and selection under uncertainty. Environmental Modelling & Software, 24(8): 901-916.
Bárdossy, A., Das, T., 2008. Influence of rainfall observation network on model calibration and application. Hydrology and Earth System Sciences, 12(1): 77-89.
Beck, M.B., 1987. Water-Quality Modelling - A review of the analysis of uncertainty. Water Resources Research, 23(8): 1393-1442.
Ben-Haim, Y., 2006. Info-Gap decision theory: Decisions under severe uncertainty. Elsevier, Amsterdam.
Bende-Michl, U., Hairsine, P.B., 2010. A systematic approach to choosing an automated nutrient analyser for river monitoring. Journal of Environmental Monitoring, 12(1): 127-134.

Berenguer, M., Zawadzki, I., 2008. A study of the error covariance matrix of radar rainfall estimates in stratiform rain. Weather and Forecasting, 23(6): 1085-1101.
Beven, K., Freer, J., 2001. A dynamic TOPMODEL. Hydrological Processes, 15(10): 1993-2011.
Beven, K., Page, T., McGechan, M., 2006. Uncertainty estimation in phosphorus models. In: Radcliffe, D.E., Cabrera, M.L. (Eds.), Modeling phosphorus in the environment. CRC Press, Boca Raton, pp. 131-160.
Beven, K., Smith, P., Freer, J., 2007. Comment on "Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini. Journal of Hydrology, 338(3-4): 315-318.
Beven, K.J., 2006. A manifesto for the equifinality thesis. Journal of Hydrology, 320(1-2): 18-36.
Beven, K.J., 2009. Comment on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" by Jasper A. Vrugt, Cajo J. F. ter Braak, Hoshin V. Gupta and Bruce A. Robinson. Stochastic Environmental Research and Risk Assessment, 23(7): 1059-1060.
Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? Journal of Hydrology, 354(1-4): 15-32.
Beven, K., Smith, P. J., and Wood, A., 2011. On the colour and spin of epistemic error (and what we might do about it), Hydrol. Earth Syst. Sci., 15, 3123-3133, doi:10.5194/hess-15-3123-2011
Beven, K.J. and Westerberg, I.K., 2011. On red herrings and real herrings: disinformation and information in hydrological inference. Hydrological Processes, 25(10): 1676-1680.
Bilotta, G.S., Krueger, T., Brazier, R.E., Butler, P., Freer, J., Hawkins, J.M.B., Haygarth, P.M., Macleod, C.J.A., Quinton, J.N., 2010. Assessing catchment-scale erosion and yields of suspended solids from improved temperate grassland. Journal of Environmental Monitoring, 12(3): 731-739.
Birkel, C., Tetzlaff, D., Dunn, S.M., Soulsby, C., 2010. Towards a simple dynamic process conceptualization in rainfall-runoff models using multi-criteria calibration and tracers in temperate, upland catchments. Hydrological Processes, 24(3): 260-275.
Birkinshaw, S. J., O'Donnell, G.M., Moore, P., Kilsby, C.G., Fowler, H.J., Berry, P.A.M., 2010. Using satellite altimetry data to augment flow estimation techniques on the Mekong River, Hydrological Processes, 24(26), 3811-3825, 10.1002/hyp.7811.
Birks, S.J., Gibson, J.J., 2009. Isotope hydrology research in Canada, 2003-2007. Canadian Water Resources Journal, 34(2): 163-176.
Bjerklie, D.M., 2007. Estimating the bankfull velocity and discharge for rivers using remotely sensed river morphology information. Journal of Hydrology, 341(3-4): 144-155.
Bjerklie, D.M., Dingman, S.L., Vorosmarty, C.J., Bolster, C.H., Congalton, R.G., 2003. Evaluating the potential for measuring river discharge from space. Journal of Hydrology, 278(1-4): 17-38.
Bjerklie, D.M., Moller, D., Smith, L.C., Dingman, S.L., 2005. Estimating discharge in rivers using remotely sensed hydraulic information. Journal of Hydrology, 309(1-4): 191-209.

Bogardi, J.J., Kundzewicz, Z.W., 1996. Introduction. In: Bogardi, J.J., Kundzewicz, Z.W. (Eds.), Risk, reliability, uncertainty, and robustness of water resources systems. Cambridge University Press, Cambridge, pp. 1-3.

Bowes, M. J., Smith, J.T., Neal, C., 2009. The value of high-resolution nutrient monitoring: A case study of the River Frome, Dorset, UK. Journal of Hydrology, 378(1-2): 82-96.

Bras, R.L., Rodriguez-Iturbe, I., 1976. Network design for the estimation of the areal means of rainfall events. Water Resources Research, 12(6): 1185-1195.

Brown, J.D., 2004. Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. Transactions of the Institute of British Geographers, 29(3): 367-381.

Brown, J.D., Heuvelink, G.B.M., 2007. The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. Computers & Geosciences, 33(2): 172-190.

Bulygina, N., McIntyre, N., Wheater, H., 2009. Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis. Hydrology and Earth System Sciences, 13(6): 893-904.

Burkham, D.E., Dawdy, D.R., 1970. Error analysis of streamflow data for an alluvial stream. Geological Survey Professional Paper 655-C. United States Government Printing Office, Washington D.C.

Burt, T.P., Howden, N.J.K., Worrall, F., Whelan, M.J., 2010. Long-term monitoring of river water nitrate: how much data do we need? Journal of Environmental Monitoring, 12(1): 71-79.

Carter, R.W., Anderson, I.E., 1963. Accuracy of current meter measurements. Journal of the Hydraulics Division, 89(4): 105-115.

Castilla, G., Hay, G.J., 2007. Uncertainties in land use data. Hydrology and Earth System Sciences, 11(6): 1857-1868.

Ciach, G.J., 2003. Local random errors in tipping-bucket rain gauge measurements. Journal of Atmospheric and Oceanic Technology, 20(5): 752-759.

Ciach, G.J., Krajewski, W.F., Villarini, G., 2007. Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. Journal of Hydrometeorology, 8(6): 1325-1347.

Clark, M., McMillan, H., Collins, D., Kavetski, D., Woods, R., 2011. Hydrological field data from a modeller's perspective: Part 2. Process-based evaluation of model hypotheses. Hydrological Processes, 25(4): 523–543.

Clark, M.P., Slater, A.G., 2006. Probabilistic quantitative precipitation estimation in complex terrain. Journal of Hydrometeorology, 7(1): 3-22.

Clarke, R.T., 1999. Uncertainty in the estimation of mean annual flood due to rating-curve indefinition. Journal of Hydrology, 222(1-4): 185-190.

Clarke, R.T., Mendiondo, E.M., Brusa, L.C., 2000. Uncertainties in mean discharges from two large South American rivers due to rating curve variability. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 45(2): 221-236.

Clesceri, L.S., Greenberg, A.E., Eaton, A.D., (Editors), 1998. Standard methods for the examination of water & wastewater. American Public Health Association, American Water Works Association and Water Environment Federation. 20th edition.

Cloke, H.L., Anderson, M.G., McDonnell, J.J., Renaud, J.P., 2006. Using numerical modelling to evaluate the capillary fringe groundwater ridging hypothesis of streamflow generation. Journal of Hydrology, 316(1-4): 141-162.

Costanza, R., Funtowicz, S. and Ravetz, J., 1992. Assessing and communicating data quality in policy-relevant research. Environmental Management, 16(1): 121-131.

Costa, J.E., Cheng, R.T., Haeni, F.P., Melcher, N., Spicer, K.R., Hayes, E., Plant, W., Hayes, K., Teague, C., Barrick, D., 2006. Use of radars to monitor stream discharge by noncontact methods. Water Resources Research, 42(7): W07422.

Di Baldassarre, G., Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. Hydrology and Earth System Sciences, 13(6): 913-921.

Dolan, D. M., Yui, A.K., Geist, R.D., 1981. Evaluation of river load estimation methods for total phosphorus. Journal of Great Lakes Research, 7(3): 207-214.

Donohue, I., Irvine, K., 2008. Quantifying variability within water samples: The need for adequate subsampling. Water Research, 42(1-2): 476-482.

Dottori, F., Martina, M.L.V., Todini, E., 2009. A dynamic rating curve approach to indirect discharge measurement. Hydrology and Earth System Sciences, 13(6): 847-863.

Dottori, F., Todini, E., 2010. Reply to Comment on 'A dynamic rating curve approach to indirect discharge measurement by Dottori et al. (2009)' by Koussis (2009). Hydrology and Earth System Sciences, 14(6): 1099-1107.

Dunn, S.M., Freer, J., Weiler, M., Kirkby, M.J., Seibert, J., Quinn, P.F., Lischeid, G., Tetzlaff, D., Soulsby, C., 2008. Conceptualization in catchment modelling: simply learning? Hydrological Processes, 22(13): 2389-2393.

Dunn, S.M., McDonnell, J.J., Vache, K.B., 2007. Factors influencing the residence time of catchment waters: A virtual experiment approach. Water Resources Research, 43(6): W06408.

Dunne, T., Black, R.D., 1970. Partial area contributions to storm runoff in a small New England watershed. Water Resources Research, 6: 1296-1311.

Dymond, J.R., Christian, R., 1982. Accuracy of discharge determined from a rating curve. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 4(12): 493-504.

Eder, A., Strauss, P., Krueger, T. and Quinton, J.N., 2010. Comparative calculation of suspended sediment loads with respect to hysteresis effects (in the Petzenkirchen catchment, Austria). Journal of Hydrology, 389(1-2): 168-176.

Essery, C.I., Wilcock, D.N., 1991. The variation in rainfall catch from standard UK Meteorological-Office rain-gages - A 12 year case-study. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 36(1): 23-34.

Evans, J.G., Wass, P.D., Hodgson, P., 1997. Integrated continuous water quality monitoring for the LOIS river programme. Science of the Total Environment, 194: 111-118.

Fekete, B. M., Vorosmarty, C. J. Grabs, W., 2002. High-resolution fields of global runoff combining observed river discharge and simulated water balances, Glob. Biogeochem. Cycle, 16(3), 1042. 10.1029/1999gb001254.

Fekete, B.M., Vorosmarty, C.J., Roads, J.O., Willmott, C.J., 2004. Uncertainties in precipitation and their impacts on runoff estimates. Journal of Climate, 17(2): 294-304.

Fenicia, F., McDonnell, J.J., Savenije, H.H.G., 2008a. Learning from model improvement: On the contribution of complementary data to process understanding. Water Resources Research, 44(6): W06419.

Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2008b. Understanding catchment behavior through stepwise model concept improvement. Water Resources Research, 44(1): W01402.

Foster, I.D.L., Millington, R., Grew, R.G., 1992. The impact of particle-size controls on stream turbidity measurement - some implications for suspended sediment yield estimation. Erosion and Sediment Transport Monitoring Programmes in River Basins, 210: 51-62.

Freer, J., McMillan, H., McDonnell, J.J., Beven, K.J., 2004. Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. Journal of Hydrology, 291(3-4): 254-277.

Gebremichael, M., Liao, G.-Y., Yan, J., 2011. Nonparametric error model for a high resolution satellite rainfall product. Water Resources Research, 47(7): W07504.

Genereux, D.P., Wood, S.J. and Pringle, C.M., 2002. Chemical tracing of interbasin groundwater transfer in the lowland rainforest of Costa Rica. Journal of Hydrology, 258(1-4): 163-178.

Gentry, L.E., David, M.B., Royer, T.V., Mitchell, C.A., Starks, K.M., 2007. Phosphorus transport pathways to streams in tile-drained agricultural watersheds. Journal of Environmental Quality, 36(2): 408-415.

Gippel, C.J., 1995. Potential of turbidity monitoring for measuring the transport of suspended-solids in streams. Hydrological Processes, 9(1): 83-97.

Goodison, B.E., Louie, P.Y.T., Yang, D., 1998. WMO Solid precipitation measurement comparison - Final Report. WMO/TD - No. 872 Instruments and Observing Methods Series. pp 318.

Goodrich, D.C., Faures, J.M., Woolhiser, D.A., Lane, L.J., Sorooshian, S., 1995. Measurement and analysis of small-scale convective storm rainfall variability. Journal of Hydrology, 173(1-4): 283-308.

Gordon, J. D., Newland, C.A., Gagliardi, S.T., 2000. Laboratory performance in the sediment laboratory quality-assurance project, 1996-98. USGS Water Resources Investigations Report 99-4184. Washington, D.C.: USGS.

Goswami, U., Sarma, J.N., Patgiri, A.D., 1999. River channel changes of the Subansiri in Assam, India. Geomorphology, 30(3): 227-244.

Goetzinger, J., Bárdossy, A., 2008. Generic error model for calibration and uncertainty estimation of hydrological models. Water Resources Research, 44: W00B07.

Graham, C.B., van Verseveld, W., Barnard, H.R., McDonnell, J.J., 2010. Estimating the deep seepage component of the hillslope and catchment water balance within a measurement uncertainty framework. Hydrological Processes, 24(25): 3878–3893.

Gupta, H.V., Wagener, T., Liu, Y.Q., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes, 22(18): 3802-3813.

Gupta, V.K., Sorooshian, S., 1985. The relationship between data and the precision of parameter estimates of hydrologic-models. Journal of Hydrology, 81(1-2): 57-77.

Haan, C.T., 1995. Fate and transport of phosphorus in the Lake Okeechobee Basin, Florida. Ecological Engineering, 5(2-3): 331-339.

Hall, J.W., 2003. Handling uncertainty in the hydroinformatic process. Journal of Hydroinformatics, 5(4): 215-232.

Hamilton, A.S., Moore, R.D., 2012. Quantifying Uncertainty in Streamflow Records. Canadian Water Resources Journal, 37(1): 3-21.

Hanafi, S., Grace, M., Webb, J.A., Hart, B., 2007. Uncertainty in nutrient spiraling: Sensitivity of spiraling indices to small errors in measured nutrient concentration. Ecosystems, 10(3): 477-487.
Harmel, R.D., Cooper, R.J., Slade, R.M., Haney, R.L., Arnold, J.G., 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of the ASABE, 49(3): 689-701.
Harmel, R.D., King, K.W., 2005. Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds. Transactions of the ASAE, 48(5): 1713-1721.
Harmel, R.D., Smith, D.R., King, K.W., Slade, R.M., 2009. Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling applications. Environmental Modelling & Software, 24(7): 832-842.
Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. Journal of Hydrology, 337(3-4): 326-336.
Hauet, A., Creutin, J.D., Belleudy, P., 2008. Sensitivity study of large-scale particle image velocimetry measurement of river discharge using numerical simulation. Journal of Hydrology, 349(1-2): 178-190.
Heberden, W., 1769. Of the different quantities of rain, which appear to fall, at different heights, over the same spot of ground. Philosophical Transactions of the Royal Society, 59: 359-262.
Heistermann, M., Kneis, D., 2011. Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling. Water Resources Research, 47(6): W06514.
Herschy, R.W., 1998. Hydrometry : Principles and practices. Wiley, Chichester.
Herschy, R.W., 2002. The uncertainty in a current meter measurement. Flow Measurement and Instrumentation, 13(5-6): 281-284.
Horowitz, A.J., 2003. An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. Hydrological Processes, 17(17): 3387-3409.
Horowitz, A.J., Rinella, F.A., Lamothe, P., Miller, T.L., Edwards, T.K., Roche, R.L., Rickert, D.A., 1990. Variations in suspended sediment and associated trace-element concentrations in selected riverine cross-sections. Environmental Science & Technology, 24(9): 1313-1320.
Horton, R.E., 1923. Accuracy of Areal Rainfall Estimates. Monthly Weather Review, 51: 348-353.
Hossain, F., Anagnostou, E.N., 2006. Assessment of a multidimensional satellite rainfall error model for ensemble generation of satellite rainfall data. IEEE Geoscience and Remote Sensing Letters, 3(3): 419-423.
Huard, D., Mailhot, A., 2006. A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc". Water Resources Research, 42(7): W07416.
Huard, D., Mailhot, A., 2008. Calibration of hydrological model GR2M using Bayesian uncertainty analysis. Water Resources Research, 44(2): W02424.
Hudson R, Fraser J., 2002. Alternative methods of flow rating in small coastal streams. Forest Research Extension Note EN-014 (Hydrology). Vancouver Forest Region.
Hutchinson, P., 1969. A note on random rain-gauge errors. Journal of Hydrology (NZ), 8(1): 8-10.

Ibbitt, R.P., 1972. Effects of random data errors on parameter values for a conceptual model. Water Resources Research, 8(1): 70-78.

ISO, 1997. Velocity area methods. International Standards Organization, Geneva.

Jalbert, J., Mathevet, T. and Favre, A.C., 2011. Temporal uncertainty estimation of discharges from rating curves using a variographic analysis. Journal of Hydrology, 397(1-2): 83-92.

Jodeau, M., Hauet, A., Paquier, A., Le Coz, J., Dramais, G., 2008. Application and evaluation of LS-PIV technique for the monitoring of river surface velocities in high flow conditions. Flow Measurement and Instrumentation, 19(2): 117-127.

Johnes, P.J., 2007. Uncertainties in annual riverine phosphorus load estimation: Impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology, 332(1-2): 241-258.

Jordan, P., Cassidy, R., 2011. Technical Note: Assessing a 24/7 solution for monitoring water quality loads in small river catchments. Hydrology and Earth System Sciences, 15(10): 3093-3100.

Jordan, P., Arnscheidt, J., McGrogan, H., McCormick, S., 2005. High-resolution phosphorus transfers at the catchment scale: the hidden importance of non-storm transfers. Hydrology and Earth System Sciences, 9(6): 685-691.

Kavetski, D., Franks, S., Kuczera, G., 2003. Confronting input uncertainty in environmental modelling. In: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Calibration of watershed models. Water Science and Applications Series. AGU, pp. 49-68.

Kavetski, D., Kuczera, G., Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research, 42(3): W03407.

Kavetski, D., Kuczera, G., Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resources Research, 42(3): W03408.

Keener, V.W., Ingram, K.T., Jacobson, B., Jones, J.W., 2007. Effects of El-Nino / Southern Oscillation on simulated phosphorus loading in South Florida. Trans. ASABE 50 (6), 2081–2089.

Kennard, M.J., Mackay, S.J., Pusey, B.J., Olden, J.D., Marsh, N., 2010. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. River Research and Applications, 26(2): 137-156.

Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. Journal of the Royal Statistical Society Series B-Statistical Methodology, 63(3): 425-450.

Khadam, I.M., Kaluarachchi, J.J., 2004. Use of soft information to describe the relative uncertainty of calibration data in hydrologic models. Water Resources Research, 40(11): W11505.

Kim, Y., Muste, M., Hauet, A., Krajewski, W.F., Kruger, A., Bradley, A., 2008. Stream discharge using mobile large-scale particle image velocimetry: A proof of concept. Water Resources Research, 44(9): W09502.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3): W03S04.

Kirstetter, P.E., Delrieu, G., Boudevillain, B., Obled, C., 2010. Toward an error model for radar quantitative precipitation estimation in the Cevennes-Vivarais region, France. Journal of Hydrology, 394(1-2): 28-41.

Klemeš, V., 1983. Conceptualisation and scale in hydrology. Journal of Hydrology, 65(1-3): 1-23.

Klemeš, V., 1996. Risk analysis: The unbearable cleverness of bluffing. In: Bogardi, J.J., Kundzewicz, Z.W. (Eds.), Risk, reliability, uncertainty, and robustness of water resources systems. Cambridge University Press, Cambridge, UK, pp. 22-29.

Kotlash, A.R., Chessman, B.C., 1998. Effects of water sample preservation and storage on nitrogen and phosphorus determinations: Implications for the use of automated sampling equipment. Water Research, 32(12): 3731-3737.

Koussis, A.D., 2010. Comment on 'A praxis-oriented perspective of streamflow inference from stage observations - the method of Dottori et al. (2009) and the alternative of the Jones Formula, with the kinematic wave celerity computed on the looped rating curve' by Koussis (2009). Hydrology and Earth System Sciences, 14(6): 1093-1097.

Krajewski, W.F., Ciach, G.J., Habib, E., 2003. An analysis of small-scale rainfall variability in different climatic regimes. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 48(2): 151-162.

Kronvang, B., Bruhn, A.J., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrological Processes, 10(11): 1483-1501.

Krueger, T., Freer, J., Quinton, J.N., Macleod, C.J.A., 2007. Processes affecting transfer of sediment and colloids, with associated phosphorus, from intensively farmed grasslands: A critical note on modelling of phosphorus transfers. Hydrological Processes, 21(4): 557-562.

Krueger, T., Freer, J., Quinton, J.N., Macleod, C.J.A., Bilotta, G.S., Brazier, R.E., Butler, P., Haygarth, P.M., 2010a. Ensemble evaluation of hydrological model hypotheses. Water Resources Research, 46: W07516.

Krueger, T., Freer, J., Quinton, J.N., Macleod, C.J.A., Bilotta, G.S., Brazier, R.E., Hawkins, J.M.B., Haygarth, P.M., 2010b. Hydrological model hypothesis testing using imprecise spatial flux measurements. In: N.J. Tate and P.F. Fisher (Editors), Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. University of Leicester, Leicester, pp. 145-148.

Krueger, T., Quinton, J.N., Freer, J., Macleod, C.J.A., Bilotta, G.S., Brazier, R.E., Butler, P., Haygarth, P.M., 2009. Uncertainties in data and models to describe event dynamics of agricultural sediment and phosphorus transfer. Journal of Environmental Quality, 38(3): 1137-1148.

Krueger, T., Quinton, J.N., Freer, J., Macleod, C.J.A., Bilotta, G.S., Brazier, R.E., Hawkins, J.M.B., Haygarth, P.M., in press. Comparing empirical models for sediment and phosphorus transfer from soils to water at field and catchment scale under data uncertainty. European Journal of Soil Science.

Kuczera, G., 1996. Correlated rating curve error in flood frequency inference. Water Resources Research, 32(7): 2119-2127.

Kuczera, G., Renard, B., Thyer, M., Kavetski, D., 2010. There are no hydrological monsters, just models and observations with large uncertainties! Hydrol. Sci. J. 55(6), 980–991.

Larson, L.W., Peck, E.L., 1974. Accuracy of precipitation measurements for hydrologic models. Water Resources Research, 10(4): 857-863.

Lazzarotto, P., Prasuhn, V., Butscher, E., Crespi, C., Fluehler, H., Stamm, C., 2005. Phosphorus export dynamics from two Swiss grassland catchments. Journal of Hydrology, 304(1-4): 139-150.
L'Ecuyer, T. S., and G. L. Stephens, 2002. An uncertainty model for Bayesian Monte Carlo retrieval algorithms: Application to the TRMM observing system. Quart. J. Roy. Meteor. Soc.,128, 1713–1737.
Legates, D.R., Willmott, C.J., 1990. Mean seasonal and spatial variability in gauge-corrected, global precipitation. International Journal of Climatology, 10(2): 111-127.
Leonard, J., Mietton, M., Najib, H., Gourbesville, P., 2000. Rating curve modelling with Manning's equation to manage instability and improve extrapolation. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 45(5): 739-750.
Lischeid, G., 2008. Combining hydrometric and hydrochemical data sets for investigating runoff generation processes: Tautologies, inconsistencies and possible explanations. Geography Compass, 2(1): 255-280.
Liu, Y.L., Freer, J., Beven, K., Matgen, P., 2009. Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. Journal of Hydrology, 367(1-2): 93-103.
Llasat, M.C., Snyder, R.L., 1998. Data error effects on net radiation and evapotranspiration estimation. Agricultural and Forest Meteorology, 91(3-4): 209-221.
Lovell, B., McKelvie, I.D., Nash, D., 2001. Sampling design for total and filterable reactive phosphorus monitoring in a lowland stream: considerations of spatial variability, measurement uncertainty and statistical power. Journal of Environmental Monitoring, 3(5): 463-468.
Mahmoud, M., Liu, Y.Q., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., White, D., Winter, L., 2009. A formal framework for scenario development in support of environmental decision-making. Environmental Modelling & Software, 24(7): 798-808.
Malcolm, I.A., Soulsby, C., Youngson, A.F., 2006. High-frequency logging technologies reveal state-dependent hyporheic process dynamics: implications for hydroecological studies. Hydrological Processes, 20(3): 615-622.
Mandapaka, P.V., Krajewski, W.F., Ciach, G.J., Villarini, G., Smith, J.A., 2009. Estimation of radar-rainfall error spatial correlation. Advances in Water Resources, 32(7): 1020-1030.
Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. Journal of Hydrology, 330(1-2): 368-381.
Mantovan, P., Todini, E., Martina, M.L.V., 2007. Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology". Journal of Hydrology, 338(3-4): 319-324.
Magnusson, B., Ellison, S.L.R., 2008. Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. Analytical and bioanalytical chemistry. 390: 201-213.

Magnusson, B., Naykki, T., Hovind, H., Krysell, M., 2004. Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories. Nordtest Technical Report 537, Edition 2. ISSN: 0283-7234. 41 p.

Martin, G. R., Smoot, J. L., White, K. D., 1992. A comparison of surface-grab and cross-sectionally integrated stream-water-quality sampling methods. Water Environ. Res. 64(7): 866-876.

McGuire, K.J., Weiler, M., McDonnell, J.J., 2007. Integrating tracer experiments with modeling to assess runoff processes and water transit times. Advances in Water Resources, 30(4): 824-837.

McIntyre, N., Lee, H., Wheater, H., Young, A., Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. Water Resources Research, 41(12): W12434.

McIntyre, N., Marshall, M., 2008. Field verification of bed-mounted ADV meters. Proceedings of the Institution of Civil Engineers-Water Management, 161(4): 199-206.

McIntyre, N., Wheater, H.S., Lees, M.J., 2002. Estimation and propagation of parametric uncertainty in environmental models. Journal of Hydroinformatics, 4(3): 177-198.

McIntyre, N.R., Wagener, T., Wheater, H.S., Chapra, S.C., 2003. Risk-based modelling of surface water quality: a case study of the Charles River, Massachusetts. Journal of Hydrology, 274(1-4): 225-247.

McIntyre, N.R., Wheater, H.S., 2004. Calibration of an in-river phosphorus model: prior evaluation of data needs and model uncertainty. Journal of Hydrology, 290(1-2): 100-116.

McMillan, H., Clark, M., Bowden, W., Duncan, M., Woods, R., 2011a. Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure. Hydrological Processes, 25(4): 511–522.

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., Clark, M., 2010. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. Hydrological Processes, 24(10): 1270-1284.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., Woods, R., 2011b. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. Journal of Hydrology, 400(1-2): 83-94.

Melching, C.S., 1995. Reliability estimation. In: Singh, V.P. (Ed.), Computer models of watershed hydrology. Water Resources Publications, Colorado, USA, pp. 69-118.

Merz, B., Thieken, A.H., 2005. Separating natural and epistemic uncertainty in flood frequency analysis. Journal of Hydrology, 309(1-4): 114-132.

Meyer, V.R., 2007. Measurement uncertainty. Journal of Chromatography A, 1158(1-2): 15-24.

Mirza, M. M. Q., 2003. The choice of stage-discharge relationship for the Ganges and Brahmaputra rivers in Bangladesh, Nordic Hydrology, 34(4), 321-342.

Moatar, F., Person, G., Meybeck, M., Coynel, A., Etcheber, H., Crouzet, P., 2006. The influence of contrasting suspended particulate matter transport regimes on the bias and precision of flux estimates. Science of the Total Environment, 370(2-3): 515-531.

Molini, A., Lanza, L.G., La Barbera, P., 2005a. The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications, Hydrological Processes, 19(5), 1073-1088.

Molini, A., Lanza, L.G., La Barbera, P., 2005b. Improving the accuracy of tipping-bucket rain records using disaggregation techniques, Atmospheric Research, 77(1-4), 203-217.

Montanari, A., 2005. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. Water Resources Research, 41(8): W08406.

Montanari, A., 2007. What do we mean by 'uncertainty"? The need for a consistent wording about uncertainty assessment in hydrology. Hydrological Processes, 21(6): 841-845.

Morrissey, M.L., Maliekal, J.A., Greene, J.S., Wang, J., 1995. The uncertainty of simple spatial averages using rain gauge networks. Water Resources Research, 31(8): 2011-2017.

Mosley, M. P., McKerchar, A.I., 1993. Streamflow. In: Handbook of Hydrology, edited by D. R. Maidment, pp. 8.1-8.39, McGraw-Hill Inc, New York.

Moulin, L., Gaume, E., Obled, C., 2009. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. Hydrology and Earth System Sciences, 13(2): 99-114.

Moyeed, R.A., Clarke, R.T., 2005. The use of Bayesian methods for fitting rating curves, with case studies. Advances in Water Resources, 28(8): 807-818.

Mueller, C.C., Kidder, E.H., 1972. Rain gage catch variation due to air-flow disturbances around a standard rain gage. Water Resources Research, 8(4): 1077-1082.

Mueller, D.S., 2003. Field evaluation of boat-mounted acoustic Doppler instruments used to measure streamflow. Proceedings of the IEEE/OES Seventh Working Conference on Current Measurement Technology. IEEE, New York, 30-34 pp.

Muste, M., Fujita, I., Hauet, A., 2008. Large-scale particle image velocimetry for measurements in riverine environments. Water Resources Research, 44: W00D19.

Neff, E.L., 1977. How much rain does a rain gauge gauge? Journal of Hydrology, 35: 213-220.

Negrel, J., Kosuth, P., Bercher, N., 2011. Estimating river discharge from earth observation measurements of river surface hydraulic variables. Hydrology and Earth System Sciences, 15(6): 2049-2058.

Nilsson, B., Hojberg, A.L., Refsgaard, J.C., Troldborg, L., 2007. Uncertainty in geological and hydrogeological data. Hydrology and Earth System Sciences, 11(5): 1551-1561.

Oberg, K., Mueller, D.S., 2007. Validation of streamflow measurements made with acoustic Doppler current profilers. Journal of Hydraulic Engineering-ASCE, 133(12): 1421-1432.

Owens, P.N., Deeks, L.K., Wood, G.A., Betson, M.J., Lord, E.I., Davison, P.S., 2008. Variations in the depth distribution of phosphorus in soil profiles and implications for model-based catchment-scale predictions of phosphorus delivery to surface waters. Journal of Hydrology, 350(3-4): 317-328.

Page, T., Beven, K.J., Freer, J., Neal, C., 2007. Modelling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty). Hydrological Processes, 21(3): 292-307.

Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research, 42(5): W05302.

Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J., de Roo, A.P.J., 2005. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). Hydrology and Earth System Sciences, 9(4): 381-393.

Pappenberger, F., Matgen, P., Beven, K.J., Henry, J.B., Pfister, L., Fraipont de, P., 2006. Influence of uncertain boundary conditions and model structure on flood inundation predictions. Advances in Water Resources, 29(10): 1430-1449.

Pappenberger, F., Cloke, H. L., Balsamo, G., Ngo-Duc, T., Oki, T., 2010. Global runoff routing with the hydrological component of the ECMWF NWP system, International Journal of Climatology, 30(14), 2155-2174, 10.1002/joc.2028.

Pelletier, P.M.., 1988. Uncertainties in the determination of river discharge: A literature review. Canadian Journal of Civil Engineering, 15: 834-850.

Pelletier, P. M., 1989. Uncertainties in streamflow measurement under winter ice conditions a case study: The Red River at Emerson, Manitoba, Canada, Water Resour. Res., 25(8), 1857–1867, doi:10.1029/WR025i008p01857.

Pelletier, P.M., 1990. A review of techniques used by Canada and other northern countries for measurement and computation of streamflow under ice conditions. Nordic Hydrology, 21(4-5): 317-340.

Perumal, M., Shrestha, K.B., Chaube, U.C., 2004. Reproduction of hysteresis in rating curves, Journal of Hydraulic Engineering-Asce, 130(9), 870-878, 10.1061/(asce)0733-9429(2004)130:9(870).

Petersen-Øverleir, A., Reitan, T., 2005. Uncertainty in flood discharges from urban and small rural catchments due to inaccurate head measurement. Nordic Hydrology, 36(3): 245-257.

Petersen-Øverleir, A., Reitan, T., 2009. Bayesian analysis of stage-fall-discharge models for gauging stations affected by variable backwater. Hydrological Processes, 23(21): 3057-3074.

Petersen-Øverleir, A., Soot, A. and Reitan, T., 2009. Bayesian Rating Curve Inference as a Streamflow Data Quality Assessment Tool. Water Resources Management, 23(9): 1835-1842.

Pfister, L., McDonnell, J.J., Wrede, S., Hlúbiková, D., Matgen, P., Fenicia, F., Ector, L., Hoffmann, L., 2009. The rivers are alive: On the potential for diatoms as a tracer of water source and hydrological connectivity. Hydrological Processes, 23(19): 2841-2845.

Phillips, J.M., Webb, B.W., Walling, D.E., Leeks, G.J.L., 1999. Estimating the suspended sediment loads of rivers in the LOIS study area using infrequent samples. Hydrological Processes, 13(7): 1035-1050.

Plate, E., 1995. Stochastic approach to non-point pollution of surface waters. In: Kundzewicz, Z.W. (Ed.), New uncertainty concepts in hydrology and water resources. Cambridge University Press, Cambridge, UK, pp. 273-283.

Quinton, J.N., Krueger, T., Freer, J., Bilotta, G.S., Brazier, R.E., 2011. A case study of uncertainty: Applying GLUE to EUROSEM. In: Morgan, R.P.C., Nearing, M.A. (Eds.), Handbook of erosion modelling. Blackwell Publishing Ltd, Chichester, pp. 80-97.

Radcliffe, D.E., Freer, J., Schoumans, O.F., 2009. Diffuse phosphorus models in the U.S. and Europe: their usages, scales, and uncertainties. Journal of Environmental Quality, 38(1-12): 1956-1967.

Reckhow, K.H., Chapra, S.C., 1979. A note on error analysis for a phosphorus retention model. Water Resources Research, 15(6): 1643-1646.

Refsgaard, J. C., Nilsson, B., Brown, J., Klauer, B., Moore, R., Bech, T., Vurro, M., Blind, M., Castilla, G., Tsanis, I., Biza, P., 2005. Harmonised techniques and representative river basin data for assessment and use of uncertainty information in integrated water management (HarmoniRiB), Environmental Science & Policy, 8 (3): 267-277

Refsgaard, J.C., van der Keur, P., Nilsson, B., Mueller-Wohlfeil, D.I., Brown, J., 2006. Uncertainties in river basin data at various support scales - Example from Odense Pilot River Basin. Hydrology Earth System Sciences Discussions, 3(4): 1943-1985.

Refsgaard, J.C., van der Sluijs, J.P., Hojberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process - A framework and guidance. Environmental Modelling & Software, 22(11): 1543-1556.

Reichert, P., Mieleitner, J., 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resources Research, 45: W10402.

Reitan, T., Petersen-Øverleir, A., 2006. Existence of the frequentistic estimate for power-law regression with a location parameter, with applications for making discharge rating curves. Stochastic Environmental Research and Risk Assessment, 20(6): 445-453.

Reitan, T., Petersen-Øverleir, A., 2008. Bayesian power-law regression with a location parameter, with applications for construction of discharge rating curves. Stochastic Environmental Research and Risk Assessment, 22(3): 351-365.

Reitan, T., Petersen-Øverleir, A., 2009. Bayesian methods for estimating multi-segment discharge rating curves. Stochastic Environmental Research and Risk Assessment, 23(5): 627-642.

Reitan, T., Petersen-Øverleir, A., 2011. Dynamic rating curve assessment in unstable rivers using Ornstein-Uhlenbeck processes. Water Resources Research, 47(2): W02524.

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., 2011. Towards a reliable decomposition of predictive uncertainty in hydrological modelling : characterizing rainfall errors using conditional simulation, Water Resources Research, 47: W11516. doi:10.1029/2011WR010643

Robards, K. McKelvie, I.D., Benson, R.L., Worsfold, P.J., Blundell, N.J., Casey, H., 1994. Determination of carbon, phosphorus, nitrogen and silicon species in waters. Analytica Chimica Acta, 287(3): 147-190.

Rode, M., Suhr, U., 2007. Uncertainties in selected river water quality data. Hydrology and Earth System Sciences, 11(2): 863-874.

Rodríguez-Iturbe, I., Mejía, J.M., 1974. The Design of Rainfall Networks in Time and Space. Water Resources Research, 10(4): 713-728.

Rossa, A. M., Cenzon, G., Monai, M., 2010. Quantitative comparison of radar QPE to rain gauges for the 26 September 2007 Venice Mestre flood, Nat. Hazards Earth Syst. Sci., 10, 371-377, doi:10.5194/nhess-10-371-2010

Rossa, A., Haase, G., Keil, C., Alberoni, P., Ballard, S., Bech, J., Germann, U., Pfeifer, M., Salonen, K., 2010. Propagation of uncertainty from observing systems into NWP: COST-731 Working Group 1. Atmospheric Science Letters, 11(2): 145-152.

Salles, C., Tournoud, M.G., Chu, Y., 2008. Estimating nutrient and sediment flood loads in a small Mediterranean river. Hydrological Processes, 22(2): 242-253.

Sarma, J.N., 2005. Fluvial process and morphology of the Brahmaputra River in Assam, India. Geomorphology, 70(3-4): 226-256.
Sauer, V.B., Meyer, R.W., 1992. Determination of error in individual discharge measurements, U.S. Geological Survey Open-File Report 92–144.
Schmidt, A.R., 2002. Analysis of stage-discharge relations for open-channel flows and their associated uncertainties. PhD Thesis. Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana-Champaign, Ill.
Schmidt, A.R., Yen, B.C., 2008. Theoretical development of stage-discharge ratings for subcritical open-channel flows. Journal of Hydraulic Engineering-ASCE, 134(9): 1245-1256.
Schmidt, P.J., Emelko, M.B., 2010. QMRA and decision-making: Are we handling measurement errors associated with pathogen concentration data correctly? Water Research, 45(2): 427-438.
Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research, 46: W10531.
Sefe, F. T. K., 1996. A study of the stage-discharge relationship of the Okavango River at Mohembo, Botswana, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 41(1), 97-116, 10.1080/02626669609491481.
Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resources Research, 38(11): 1241.
Seo, B.C., Krajewski, W.F., 2010. Scale dependence of radar rainfall uncertainty: Initial evaluation of NEXRAD's new super-resolution data for hydrologic applications. Journal of Hydrometeorology, 11(5): 1191-1198.
Sevruk, B., 1982. Methods of correction for systematic error in point precipitation measurement. World Meteorological Organisation, Operational Hydrology Report No. 21, WMO-No.589. Geneva, Switzerland.
Sevruk, B., 1987. Point precipitation measurements: Why are they not corrected?, Water for the future: Hydrology in perspective (Proceedings of the Rome Symposium, April 1987). IAHS Publications 164, pp. 477-486.
Sevruk, B., 1996. Adjustment of tipping-bucket precipitation gauge measurements. Atmospheric Research, 42(1-4): 237-246.
Shiklomanov, A.I., Yakovleva, T.I., Lammers, R.B., Karasev, I.P., Vörösmarty, C.J., Linder, E., 2006. Cold region river discharge uncertainty - Estimates from large Russian rivers. Journal of Hydrology, 326(1-4): 231-256.
Shimizu, Y., Giri, S., Yamaguchi, S., Nelson, J., 2009. Numerical simulation of dune-flat bed transition and stage-discharge relationship with hysteresis effect. Water Resources Research, 45: W04429.
Shrestha, R.R., Bárdossy, A., Nestmann, F., 2007. Analysis and propagation of uncertainties due to the stage-discharge relationship: A fuzzy set approach. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 52(4): 595-610.
Shrestha, R.R., Simonovic, S.P., 2010a. Fuzzy set theory based methodology for the analysis of measurement uncertainties in river discharge and stage. Canadian Journal of Civil Engineering, 37(3): 429-439.
Shrestha, R.R., Simonovic, S.P., 2010b. Fuzzy Nonlinear Regression Approach to Stage-Discharge Analyses: Case Study, Journal of Hydrologic Engineering, 15(1), 49-56, 10.1061/(asce)he.1943-5584.0000128.

Sieck, L.C., Burges, S.J., Steiner, M., 2007. Challenges in obtaining reliable measurements of point rainfall. Water Resources Research, 43(1): W01420.

Singh S. K., Bardossy A., 2012. Calibration of hydrological models on hydrologically unusual events, Advances in Water Resources 38: 81-91.

Singh, V.P., Woolhiser, D.A., 1976. Sensitivity of linear and nonlinear surface runoff models to input errors. Journal of Hydrology, 29(3-4): 243-249.

Sivakumar, B., 2004. Dominant processes concept in hydrology: Moving forward. Hydrological Processes, 18(12): 2349-2353.

Slade, R. M., 2004. General Methods, Information, and Sources for Collecting and Analyzing Water-Resources Data. CD-ROM. Copyright 2004 Raymond M. Slade, Jr.

Sorooshian, S., 1981. Parameter-estimation of rainfall-runoff models with heteroscedastic streamflow errors – The non-informative data case. Journal of Hydrology, 52(1-2): 127-138.

Sorooshian, S., Gupta, V.K., Fulton, J.L., 1983. Evaluation of maximum-likelihood parameter-estimation techniques for conceptual rainfall-runoff models – Influence of calibration data variability and length on model credibility. Water Resources Research, 19(1): 251-259.

Soulsby, C., Malcolm, R., Helliwell, R., Ferrier, R.C., Jenkins, A., 2000. Isotope hydrology of the Allt a' Mharcaidh catchment, Cairngorms, Scotland: implications for hydrological pathways and residence times. Hydrological Processes, 14(4): 747-762.

Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research, 44: W00B06.

Steiner, M., 1996. Uncertainty of estimates of monthly areal rainfall for temporally sparse remote observations. Water Resources Research 32(2): 373-388.

Steiner, M., Smith, J.A., Burges, S.J., Alonso, C.V., Darden, R.W., 1999. Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. Water Resources Research, 35(8): 2487-2503.

Stephens, G.L. and Kummerow, C.D., 2007. The remote sensing of clouds and precipitation from space: A review. Journal of the Atmospheric Sciences, 64(11): 3742-3765.

Stevens, R. J., Smith, R.V., 1978. A comparison of discrete and intensive sampling for measuring the loads of nitrogen and phosphorus in the river main, County Antrim. Water Research, 12(10): 823-830.

Storm, B., Jensen, K.H., Refsgaard, J.C., 1989. Estimation of Catchment Rainfall Uncertainty and its Influence on Runoff Prediction. Nordic Hydrology, 19: 77-88.

Tarboton, D. G., Maidment, D. R., Zaslavsky, I., Ames, D.P., Goodall, J., Horsburgh, J.S., 2010. CUAHSI Hydrologic Information System 2010 Status Report, Consortium of Universities for the Advancement of Hydrologic Science, Inc, 34 p.

Tarboton, D. G., Horsburgh, J. S., Maidment, D.R., 2008. CUAHSI Community Observations Data Model (ODM), Version 1.1, Design Specifications. 58 p.

Teixeira, E.C., Caliari, P.C., 2005. Estimation of the concentration of suspended solids in rivers from turbidity measurement: error assessment. In: Walling, D.E., Horowitz, A.J. (Eds.), Sediment budgets 1. IAHS Publications 291, pp. 151-160.

Tetzlaff, D., Uhlenbrook, S., Eppert, S., Soulsby, C., 2008. Does the incorporation of process conceptualization and tracer data improve the structure and performance of a simple rainfall-runoff model in a Scottish mesoscale catchment? Hydrological Processes, 22(14): 2461-2474.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. Water Resources Research, 45: W00B14.

Troutman, B.M., 1982. An analysis of input errors in precipitation-runoff models using regression with errors in the independent variables. Water Resources Research, 18(4): 947-964.

Troutman, B.M., 1983. Runoff prediction errors and bias in parameter-estimation induced by spatial variability of precipitation. Water Resources Research, 19(3): 791-810.

van Buren, M.A., Watt, W.E., Marsalek, J., 1997. Application of the log-normal and normal distributions to stormwater quality parameters. Water Research, 31(1): 95-104.

van der Keur, P., Iversen, B.V., 2006. Uncertainty in soil physical data at river basin scale - A review. Hydrology and Earth System Sciences, 10(6): 889-902.

van der Made, J.E., 1982. Determination of the accuracy of water level observations, Proceedings of the Exeter Symposium. IAHS Publications 134, pp. 172-184.

Vasiloff, S.V., Howard, K.W., Zhang, J. 2009. Difficulties with Correcting Radar Rainfall Estimates Based on Rain Gauge Data: A Case Study of Severe Weather in Montana on 16-17 June 2007. Weather and Forecasting 24(5): 1334-1344

Venetis, C., 1970. A note on the estimation of the parameters in logarithmic stage–discharge relationships with estimation of their error. Hydrological Sciences Bulletin-Bulletin Des Sciences Hydrologiques, 15(2): 105-111.

B.E. Vieux, J.M. Imgarten., 2011. On the scale-dependent propagation of hydrologic uncertainty using high-resolution X-band radar rainfall estimates. Atmospheric Research. 103: 96-105.

Villarini, G., Krajewski, W.F., 2008. Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. Advances in Water Resources, 31(7): 1015-1023.

Villarini, G., Krajewski, W.F., 2010. Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. Surveys in Geophysics, 31(1): 107-129.

Villarini, G., Krajewski, W.F., Ciach, G.J., Zimmerman, D.L., 2009. Product-error-driven generator of probable rainfall conditioned on WSR-88D precipitation estimates. Water Resources Research, 45: W01404.

Viney, N.R. and Bates, B.C., 2004. It never rains on Sunday: The prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. International Journal of Climatology, 24(9): 1171-1192.

Viviroli, D., Weingartner, R., Messerli, B., 2003. Assessing the hydrological significance of the world's mountains. Mountain Research and Development, 23(1): 32-40.

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41(1): W01017.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research, 39(8): 1201.
Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research, 43(1): W01411.
Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research, 44: W00B09.
Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2009a. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stochastic Environmental Research and Risk Assessment, 23(7): 1011-1026.
Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2009b. Response to comment by Keith Beven on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?". Stochastic Environmental Research and Risk Assessment, 23(7): 1061-1062.
Wagener, T., Freer, J.E., Zehe, E., Beven, K.J., Gupta, H.V., Bardossy, A., 2006. Towards an uncertainty framework for predictions in ungauged basins: The uncertainty working group. In: Sivapalan, M., Wagener, T., Uhlenbrook, S., Zehe, E., Lakshmi, V., Liang, X., Tachikawa, Y., Kumar, P. (Eds.), Predictions in ungauged basins (PUB): Promise and progress. IAHS Publications 303, pp. 454-462.
Wagener, T., Gupta, H., Yatheendradas, S., Goodrich, D.C., Unkrich, C.L., Schaffner, M., 2007. Understanding sources of uncertainty in flash-flood forecasting for semi-arid regions. In: Boegh E, K.H., Wagener T, Hall A, Bastidas L, Franks S, Gupta HV, Rosbjerg D and Schaake J (Ed.), Quantification and reduction of predictive uncertainty for sustainable water resources management, IAHS Publications 313, pp. 204-212.
Wagener, T., Wheater, H.S., 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. Journal of Hydrology, 320(1-2): 132-154.
Walling, D.E., Teed, A., 1971. A simple pumping sampler for research into suspended sediment transport in small catchments. Journal of Hydrology, 13: 325-337.
Walling, D.E., Webb, B.W., 1981. The reliability of suspended sediment load data, Erosion and sediment transport measurement. IAHS Publications 133, pp. 177-194.
Ward, P.R.B., 1984. Measurement of sediment yield. In: Hadley, R.F., Walling, D.E. (Eds.), Erosion and sediment yield: Some methods of measurement and modelling. Geo Books, Norwich, UK, pp. 37-70.
Wass, P.D., Leeks, G.J.L., 1999. Suspended sediment fluxes in the Humber catchment, UK. Hydrological Processes, 13(7): 935-953.
Wass, P.D., Marks, S.D., Finch, J.W., Leeks, G.J.L., Ingram, J.K., 1997. Monitoring and preliminary interpretation of in-river turbidity and remote sensed imagery for suspended sediment transport studies in the Humber catchment. Science of the Total Environment, 194: 263-283.
Wechsler, S.P., 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrology and Earth System Sciences, 11(4): 1481-1500.

Westerberg, I., Guerrero, J.L., Seibert, J., Beven, K.J., Halldin, S., 2011. Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. Hydrological Processes, 25(4): 603-613.
Whalley, N., Iredale, R.S., Clare, A.F., 2001. Reliability and uncertainty in flow measurement techniques - Some current thinking. Physics and Chemistry of the Earth Part C-Solar-Terrestial and Planetary Science, 26(10-12): 743-749.
Widen-Nilsson, E., Halldin, S. and Xu, C.Y., 2007. Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation. Journal of Hydrology, 340(1-2): 105-118.
Willems, P., 2001. Stochastic description of the rainfall input errors in lumped hydrological models. Stochastic Environmental Research and Risk Assessment, 15(2): 132-152.
WMO, 1994. Guide to Hydrological Practices, 5th Edition. World Meteorological Organization Report No. 168, pp 770.
WMO, 2008a. Abridged final report of the thirteenth session of the Commission for Hydrology: Resolution 2 (CHy XIII) Project for the assessment of the performance of flow measurement instruments and techniques, World Meteorological Organization, Geneva 4-12 November 2008. p 27-35.
WMO, 2008b. Guide to meteorological instruments and methods of observation, 7th Edition. Report WMO-8, pp 681.
WMO, 2010. Assessment of the performance of flow measurement instruments and techniques - Working website. World Meteorological Organization. http://www.wmo.int/pages/prog/hwrp/Flow/flow_tech/index.php. Accessed 22.09.2010.
Wood, S.J., Jones, D.A., Moore, R.J., 2000. Accuracy of rainfall measurement for scales of hydrological interest. Hydrology and Earth System Sciences, 4(4): 531-543.
Worsfold, P.J., Gimbert, L.J., Mankasingh, U., Omaka, O.N., Hanrahan, G., Gardolinski, P.C.F.C., Haygarth, P.M., Turner, B.L., Keith-Roach, M.J., McKelvie, I.D., 2005. Sampling, sample treatment and quality assurance issues for the determination of phosphorus species in natural waters and soils. Talanta, 66(2): 273-293.
Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources, 30(8): 1756-1774.
Yang, D.Q., Goodison, B.E., Ishida, S., Benson, C.S., 1998. Adjustment of daily precipitation data at 10 climate stations in Alaska: Application of World Meteorological Organization intercomparison results. Water Resources Research, 34(2): 241-256.
Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., Stewart, A., 2008. Understanding uncertainty in distributed flash flood forecasting for semiarid regions. Water Resources Research, 44(5): W05S19.
Young, P., 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. Hydrological Processes, 17(11): 2195-2217.
Younger, P.M., Freer, J.E., Beven, K.J., 2009. Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. Hydrological Processes, 23(14): 1988-2003.
Zappa, M., Beven, K.J., Bruen, M., Cofiño, A.S., Kok, K., Martin, E., Nurmi, P., Orfila, B., Roulin, E., Schröter, K., Seed, A., Szturc, J., Vehviläinen, B., Germann, U., Rossa, A., 2010. Propagation of uncertainty from observing

systems and NWP into hydrological models: COST-731 Working Group 2. Atmospheric Science Letters, 11(2): 83-91.

Table 1a: Typical quantitative results of rainfall uncertainty studies: Point Measurements. Bold values were used in Figure 1.

Uncertainty Type	Estimation Method	Magnitude	Location	Reference
Systematic Errors				
Wind loss curves dependent on wind speed & raindrop size	Theoretical calculation using wind velocity field from wind tunnel experiments	1 mm drops: -10 % (6 m s^{-1}), -40 % (9 m s^{-1}), -80 % (12 m s^{-1}) 2 mm drops: -10 % (9 m s^{-1}), -20 % (12 m s^{-1}) 3-5 mm drops: no effect up to 15 m s^{-1}		Mueller & Kidder (1972)
Wind loss curves	Comparison with shielded gauge	Approx. linear 1 % under-catch per 1 mph wind speed	Danville, Vermont, USA	Larson & Peck (1974); also wind loss curves for snow
Undercatch for gauge mounted at 1 m height	Comparison with pit gauge	**5-16 % average undercatch (over 53-321 events), 0-75 % per storm**	USA: Reynolds Creek, Idaho; Pullman, Washington; Sidney, Montana; Ekalaka, Montana	Neff (1977)
Loss due to wind field deformation	WMO literature survey & pit gauge comparisons	**2-10 % (rain), 10-50 % (snow)**		Sevruk (1982); extensive literature survey is still widely quoted; correction equations are given dependent on gauge type & meteorological conditions
Wetting loss		**2-15 % (summer), 1-8 % (winter)**		
Evaporation loss from open container		**0-4 %**		
Splash-in/out		**1-2 %**		
Undercatch for shielded gauge at 12 inches height & turf wall gauge	Comparison with pit gauge	**5 % (unshielded), 2 % (turf wall) annual undercatch**	County Londonderry, Ireland. Lowland, coastal, rainfall 900-1100 mm yr^{-1}.	Essery & Wilcock (1991); 1976-1988
Wind-induced error depending on wind speed, rain drop size distribution & gauge design	Comparison between exposed & pit gauges	**2–10 % (hourly data; even after popular correction algorithms)**	ARS Goodwin Creek experimental watershed, Mississippi, USA. 21.4 km^2, rainfall 1400 mm yr^{-1}, 71-128 m a.s.l.	Sieck et al. (2007)
Tipping error per 1 mm rain	Field calibration with known water delivery rate	**Up to 10 % dependent on gauge type & rain rate**		
Random Errors				
Coefficient of variation of random errors	12 co-located standard rain gauges	Approx. 5 % for single storm, independent of total storm rainfall	Mount Cargill, Dunedin, New Zealand. Exposed site	Hutchinson (1969)

			at 560 m a.s.l.	
Coefficient of variation of non-recording gauges	9 co-located recording & non-recording gauges	**4-5 % for storms >15 mm (monsoon season thunderstorms)**	USDA Walnut Gulch Experimental Watershed, Arizona, USA. 4.4 ha, semi-arid, 1250-1585 m a.s.l.	Goodrich et al. (1995)
Total error of recording gauge	Standard error between single gauges & average of 15 co-located tipping buckets	**Decreases with rain rate & accumulation time, e.g. 4.9 % (5 min) & 2.9 % (15 min) at rain rate of 10 mm h^{-1}**	USDA field station in Chickasha, Oklahoma, USA	Ciach (2003)

Table 1b: Typical quantitative results of rainfall uncertainty studies: Interpolation. Bold values were used in Figure 1.

Uncertainty Type	Estimation Method	Magnitude	Location	Reference
Rainfall variability in convective events	48 non-recording gauges on 30 m grid over 4.4 ha catchment	**4-14 % variation of mean storm rainfall over 100 m distance; -5.6 % greatest difference between areal mean & 4 co-located central gauges**	USDA Walnut Gulch Experimental Watershed, Arizona, USA. 4.4 ha, semi-arid, 1250-1585 m a.s.l.	Goodrich et al. (1995)
Standard error in single gauge measurement vs. gauge network	8 rain gauges within a 2 km^2 area	**33 % (low relief), 45 % (high relief) at 4 mm/15 min rain rate; 90% confidence bounds on the standard error, dependent on rain rate, are also given graphically**	Brue catchment, UK (135 km^2). 20-250 m a.s.l., temperate climate, orographic rainfall.	Wood et al. (2000)
	49 rain gauges in 135 km^2 area	**65 % at 4 mm/15 min rain rate; presented graphically for rain rates 0.2-8 mm/15 min and for three different gauges**		
SD of rainfall rates within 5 km^2 area for accumulation periods between 5 min and 1 hour	5 clusters, each of 12-40 rain gauges	12.2, 12.0, 16.1, 7.7 & 9.8 mm h^{-1} for 5 min totals over 57-515 days, conditioned on rain rates greater than 0.5 mm h^{-1}	Gauge clusters in Guam, Brazil, Florida, Oklahoma, Iowa	Krajewski et al. (2003); also looked at correlation statistics up to 8 km distance with significant reductions
Multiplier from 3-gauge average to areal mean rainfall	Conditional simulation using 13 raingauges to generate ensemble of spatial rainfall fields	Rainfall multipliers have mean 1.15 ± 0.03, standard deviation 0.27 ± 0.02 when accounting separately for rainfall, runoff and structural uncertainty.	Yzeron catchment (129 km^2), Rhone-Alpes region, France. 400-917 m a.s.l.. Rainfall 845 mm yr^{-1}, runoff	Renard et al. (2011)

			150 mm yr^{-1}.	

Table 1c: Typical quantitative results of rainfall uncertainty studies: Radar and Satellite. Bold values were used in Figure 1.

Uncertainty Type	**Estimation Method**	**Magnitude**	**Location**	**Reference**
Radar				
Error between radar estimate and gauge network	Radar RMSE with respect to 30 raingauges	10 % for storms >30 mm after radar bias correction using high quality rain gauge data; when all gauges were used for bias correction without prior quality control RMSE was 10-40 %	ARS Goodwin Creek experimental watershed, Mississippi, USA	Steiner et al. (1999)
Error between radar estimate and gauge network	Standard error of residuals compared with 8 rain gauges in 2 km^2 area	**50% (low relief) at 4 mm/15 min rain rate; presented graphically for rain rates 0.4-10 mm/15 min**	Brue catchment, UK (135 km^2). 20-250 m a.s.l., temperate climate, orographic rainfall.	Wood et al. (2000)
	Standard error of residuals compared with 49 rain gauges in 135 km^2 area	**55 % at 2km resolution, 60 % at 5 km resolution, for rain rate 4 mm/15 min; presented graphically for rain rates 0.2-8 mm/15 min**		
Error between radar (WSR-88D) estimate and gauge network	SD of the stochastic component of multiplicative error	**Conditioned on distance from radar, timescale of observation & season; asymptotic SD at high rainfall rates in the range 0.1-0.7, typically 0.5 for hourly data**	Oklahoma, USA. Rainfall 800 mm yr^{-1}, dominated by midlatitude convective systems.	Ciach et al. (2007)
Error between radar (S-band) estimate and gauge network	SD of residuals	**Approx. 0.3 (proportion of mean rain rate) for hourly data over 0-100 km distance from radar; values also given for 1, 2, 6, 12 hours & 0-50, 50-100, 0-100 km distances**	Cévennes-Vivarais region, France. 200 km *160 km convective and frontal rainfall.	Kirstetter et al. (2010)
Error between radar (WSR-88D) estimate and gauge network	SD of residuals (2 research gauge networks)	**0.48 (hourly, 8 km resolution), 1.07 (hourly, 1 km resolution), proportion of mean rain rate; values also given for 15 min, 1 hour at scales 0.5, 1, 2, 4, 8 km**	Iowa, USA	Seo & Krajewski (2010); raingauge networks used paired gauges at all sites
Error between radar (X-band) estimate and gauge network	Mean and SD of bias for pixel-based comparison between 2 radars and 20 gauges.	Using a Z-R relationship to estimate rainfall, the mean bias for the 2 radars was -0.24, -0.27; with SD of the relative error 0.46, 0.48.	Southwest Oklahoma, USA. Raingauges – radar distance up to 35 km. Study used 4 storm events of heavy/	Vieux and Imgarten (2011)

			broken squall lines with embedded convective cells.	
Satellite				
Bias in estimates of surface rain rate from TRMM (Tropical Rainfall Measuring Mission)	Bayesian modelling approach to estimate SD of each parameter in algorithm used to calculate surface rain rate	**SD of combined multiplicative bias in rain rate presented graphically as a function of rain rate: 40-60% at rates up to 18 mm h^{-1}, 150 % at 25 mm h^{-1},**	All oceanic pixels for 10 TRMM orbits	L'Ecuyer and Stephens (2002)
Bias of two NASA satellite products (infrared & passive microwave)	Mean & variance in multiplicative bias at hourly timesteps & 0.25º resolution compared with ground radar	**Mean multiplicative hourly bias 0.35-1.09 (with SD of 0.73-0.84) over 4-month study period.**	Oklahoma, USA. Southern Plains, 95-100°W, 34-37°N.	Hossain & Anagnostou (2006)

RMSE = root mean square error; SD = standard deviation

Table 2a: Typical quantitative results of discharge uncertainty studies: Stage Uncertainty. Bold values were used in Figure 2.

Uncertainty Type	Estimation Method	Magnitude	Location	Reference
Stage uncertainty	Comparison with neighbouring stations	SD of 25 mm	Netherlands gauging network	Van der Made (1982)
Effect of unstable bed	Expert knowledge; uncertainty for individual measurement	±10 %	Estimate for locations with shifting sand or moving dunes	Sauer & Meyer (1992)
Instrument precision	Review of previous studies; uncertainty for individual measurement	**±3-10.8 mm or ±0.1-2 %**		Quoted in Harmel et al. (2006)
Instrument precision	Expert knowledge	**Range ±10 mm; local oscillations of water surface can add additional uncertainty of ±20 mm**	Typical example of natural rivers	Dottori et al. (2009)
Instrument precision: Float in stilling well		**6 mm**		Quoted in Herschy (1998): Ackers et al. (1978)
Instrument precision: Pressure transducer		**1.4-40 mm**		Herschy (1998)
Stage uncertainty	Expert knowledge of typical uncertainties	**4 mm (high accuracy) to 15 mm (low accuracy)**	Norwegian Water Resources & Energy Directorate	Petersen-Øverleir & Reitan (2005)
Stage uncertainty	Observed fluctuation	**2-5 mm**	Rowden Experimental Research Platform (1 ha fields), Devon, UK. 250 x 37 cm weir box, stainless steel 45° V-Notch, float (Model 6541, Unidata), stilling well, ave. annual precipitation 1055 mm.	Krueger et al. (2010a)
Instrument precision	Manufacturer cited random uncertainty	**2.5 mm (Trutrack, Model PLUT-HR Water level recorder)**	Hillslope (172 m^2), WS10 catchment, HJA Experimental Forest, Oregon, USA	Graham et al. (2010)
		0.3 mm (Model 2 Stevens Instruments Position Analog Transmitter)	WS10 catchment (10.2 ha), HJA Experimental Forest, Oregon, USA. Mediterranean climate, rainfall	

			2200 mm yr^{-1}, slopes 30-45 °.	
Stage uncertainty	Nominal uncertainty	**3 mm**		Hamilton & Moore (2012)

Table 2b: Typical quantitative results of discharge uncertainty studies: Discharge Uncertainty. Bold values were used in Figure 2.

Uncertainty Type	**Estimation Method**	**Magnitude**	**Location**	**Reference**
Instantaneous Discharge Uncertainty				
Single discharge measurement uncertainty when using method of verticals with current meter	SD of relative discharge error calculated from individual uncertainty components	**2.3 % using 30 verticals with measurements at 0.2 & 0.8 depth points; other combinations also given**	Columbia River, USA (5 sites)	Carter & Anderson (1963)
Single discharge measurement uncertainty using velocity-area method	95 % confidence bounds on relative uncertainty, from literature review	4-17% for 35-5 verticals at 0.25 m s^{-1} 5-40 % for velocities 0.5-0.05 m s^{-1}.	Various	Pelletier (1988)
Single discharge measurement uncertainty under ice	Difference between USGS & Water Survey of Canada instantaneous flow measurements attributed to different setup of current meter on rod or in suspension	**2-17 %**	Red river at Emerson, Manitoba, Canada (104000 km^2). Slope 0.04-0.25 m km^{-1}, mean discharge 94.2 m^3 s^{-1}, when under ice 20 m^3 s^{-1}, drains glacial plain with moraines.	Pelletier (1989)
Combination of stage error & components of discharge error for wading or cable methods	Standard error computed by root-mean-square of component uncertainties: those derived from previous studies, manufacturer citations and expert knowledge.	**2.4 % (Good Cable)** **4.0 % (Good Wading)** **19 % (Poorest measurements)**		Sauer & Meyer (1992)
Single discharge measurement uncertainty: effect of reducing number of verticals	Halving number of verticals	**Approx. 5 % (given as graph relating to % reduction in verticals)**	23 sites in UK North-East	Whalley (2001)
Epistemic single discharge measurement uncertainty using	Combined uncertainty values from expert opinion &	**6 %**	Typical example	Herschy (2002)

current meter for velocity-area method	previous studies			
Single discharge measurement uncertainty: Salt dilution gauging	SD of instantaneous discharge measured using salt dilution, deviation from rating curve developed using both salt dilution and current metering.	**5 %**	Stephanie Creek, Vancouver Island, BC, Canada (8.6 km^2). Steep rocky creek.	Hudson & Fraser (2002)
		7.1%	Flume Creek, Sunshine Coast, BC, Canada (118 ha). Steep creek.	
		±42-84 %	South Fork catchment (780 km^2), Iowa, USA	
Single discharge measurement uncertainty	Typical bias determined from replicates	<-4 %		Hamilton & Moore (2012)
Rating Curve and Combined Uncertainty				
Random errors associated with power law rating curves	RMSE of component uncertainties	1.9 % in instantaneous or average daily discharge, 0.5 % in average monthly discharge	Mangawhero at Ore Ore, New Zealand. Mean discharge 13m^3 s^{-1}	Dymond & Christian (1982)
Deviation between theoretical & measured rating curve (with current meter)		20 % at low flows (0.2 m above station datum), 10 % at higher flows	Sprint, UK. Flat-vee crump profile weir structure.	Whalley (2001)
Deviation between theoretical rating curve accounting for non-steady flow & measured discharge (also given for empirical rating curve)	Coefficient of variation calculated from 55 discharge measurements	10 % (in-bank flows) 36% (including out-of-bank flows)	Illinois River, USA. Low gradient river, discharge 38-3480 m^3 s^{-1}, two gauge (slop-stage-discharge) rating station.	Schmidt & Yen (2008)
Total instantaneous discharge uncertainty caused by interpolation / extrapolation of rating curve, unsteady flow conditions & seasonal changes in roughness	95 % uncertainty bounds for relative error calculated through combination of three error components	**6.2 % at 1000 m^3 s^{-1} to 42.8 % at 12000 m^3 s^{-1}, average 25.6 %**	Po River, Italy (70000 km^2). Channel width 200-500 m, depth 10-15 m, slope 0.02, floodplain width 1000-3000 m.	Di Baldassarre & Montanari (2009)
Total instantaneous discharge uncertainty caused by rating curve uncertainty	Relative error compared to manual measurements	1-20 % (average 8.76 %), negatively related to stage	Hillslope (172 m^2), WS10 catchment, HJA Experimental Forest, Oregon, USA. Stilling well	Graham et al. (2010) ; values calculated from original figures

			with 30° V-Notch Weir.	
		Average 3.6 %, not related to stage	WS10 catchment (10.2 ha), HJA Experimental Forest, Oregon, USA. 90° V-Notch Weir	
Total instantaneous discharge uncertainty caused by gauging errors & rating curve form / extrapolation	Estimate of upper & lower discharge bounds for any given stage through combination of component errors	**Relative error from 100 % (low flows) to 10 % (low-mid flows) to 20 % (high flows)**	Rowden Experimental Research Platform (1 ha fields), Devon, UK. 250 x 37 cm weir box, stainless steel 45° V-Notch, bucket method & electromagnetic flowmeter (Magflo Mag 5100, Siemens), ave. annual precipitation 1055 mm.	Krueger et al. (2010a)
Total instantaneous discharge uncertainty caused by gauging error, rating curve form / extrapolation & instability of rating curve	Estimate of complete instantaneous discharge PDF for any given stage	**Relative error from 46 % (low flows) to 10 % (mid-high flows) to 15 % (flood flows), average 22 %**	Wairau River, New Zealand (3825 km^2). Elevation 0-2309 m a.s.l., braided reach, 100 m width.	McMillan et al. (2010)
Total instantaneous discharge uncertainty caused by gauging error & instability of rating curve	Estimates of upper & lower instantaneous discharge bounds for any given stage using uncertain time-varying rating curve	**Difference from constant rating curve ranged from -60 to 90 % (low flows) to ±20 % (mid-high flows); total relative discharge error -43 % to +73 %. Effect of using only 3 stage measurements / day to calculate daily discharge: ±17 %**	Choluteca River, Honduras (1766 km^2). Mountainous, 660 – 2320 m a.s.l., precipitation mainly convective.	Westerberg et al. (2011)
Time-averaged Discharge Uncertainty				
Total uncertainty of daily discharge	PDF, mean, SD	Normal, 0, 10 %	Odense basin (1190 km^2), Denmark. Low rolling hills, elevation 0-100 m a.s.l.	Refsgaard et al. (2006)
Relative uncertainty of daily & annual discharge estimates in rivers subject to icing	Statistical analysis of uncertainty in the parameters of the fitted quadratic rating	Where cross sections assumed stable: 8-25 % for low flows, 2-5 % for high flows (variation for	6 largest Eurasian Arctic Rivers (248000-2950000 km^2). Mean discharge 2200-18400 m^3 s^{-1}.	Shiklomanov et al. (2006)

	curves & ice correction coefficients	different rivers); where cross section not stable (e.g. with ice): 10-21 % with high frequency gaugings, 15-45 % under the worst conditions in the record		
Monthly discharge uncertainty	Probable error range	±42 %	Small watershed near Riesel, Texas, USA	Harmel & Smith (2007) based on Harmel et al. (2006)
Daily discharge uncertainty		±42 %; ±100-200 % for low flows; ±100 % for high flows	Reynolds Creek catchment (239 km^2), Idaho, USA	
Storm discharge uncertainty	Total probable error based on RMSE propagation method	2-19 %	Various in USA (2.2-5506 ha)	Harmel et al. (2009) based on Harmel et al. (2006)
Deep seepage uncertainty in steady state (as residual water balance component)	Relative uncertainty based on propagation of component uncertainties	57 % (under steady state); 32 % (during irrigation); 34 % (during irrigation + 5 days); 35 % (during irrigation + 10 days)	Hillslope (172 m^2), WS10 catchment, HJA Experimental Forest, Oregon, USA. Stilling well with 30° V-Notch Weir.	Graham et al. (2010) ; values calculated from original figures
		84 % (under steady state); 62 % (during irrigation); 93 % (during irrigation + 5 days); 155 % (during irrigation + 10 days)	WS10 catchment (10.2 ha), HJA Experimental Forest, Oregon, USA. 90° V-Notch Weir	
Daily discharge; effect of manual stage reading	Manually minus automatically derived discharge	Up to ±10-50 %	Lillooet River near Pemberton, British Columbia, Canada. Nivo-glacial.	Hamilton & Moore (2012)
Monthly discharge; effect of manual stage reading		Up to 5-10 %		

Table 2c: Typical quantitative results of discharge uncertainty studies: New Measurement Techniques. Bold values were used in Figure 2.

Uncertainty Type	Estimation Method	Magnitude	Location	Reference
ADCP discharge measurement uncertainty	Relative error of discharge calculated using ADCP vs. current meter and/or rating curve	**Mean relative error from multiple transects was -3 to 5 % (from meter) or -7 to 5 % (from rating) dependent on site**	USA (5 sites on Illinois, Kankakee, Mississippi and Missouri rivers). Depths 1.1-3.8 m, widths 33-527 m, velocities 0.4-1.3 m s^{-1}.	Mueller (2003)
	Relative error of discharge	SD of relative error 5.8 %;	Multi-location field sites (including	Oberg & Mueller (2007)

	calculated using ADCP vs. multiple concurrent current meters	distributions given from large set of test cases, plus results for alternative measurement set-ups	USA, Canada, Sweden, Netherlands) plus laboratory testing	
ADV velocity measurement uncertainty, with & without calibration	Relative error of discharge calculated using ADV velocity (20 min average) vs. impellor velocity (60 s period per sample)	**Flow estimates were within 20 % of the current-metered flow for 93 % of samples after calibration (68 % before calibration)**	Pontbren, Wales, UK, 5 concrete-lined sections. 3 circular: diameter 0.6-1.6 m, depth 0-0.71 m, velocity 0-3.0 m s^{-1}. 2 rectangular: width 3.17, 4.17 m; depth 0-0.67 m, velocity 0-3.9 m s^{-1}.	McIntyre & Marshall (2008)
Mobile LSPIV instantaneous velocity & discharge measurement uncertainty	Relative error from theoretical velocity field based on 27 error sources; case study comparison with rating curve & ADCP methods	**Theoretical errors in velocity from 10-35 % at 95 % confidence level; case study gave discharge error at 2 % compared to rating curve & 5.5 % compared to ADCP**	Analysis of typical conditions. Case study at Clear Creek near Coralville, Iowa, USA. 20 m wide, 0.7 m deep, stage 1.2 and velocity 5.2 m s^{-1} during study.	Kim et al. (2008)
Simulated LSPIV measurements against theoretical true values	Error variance obtained via linear regression of simulated vs. true values	**5 % under normal conditions, increasing to 17 % with a high tilt angle (70º)**	Numerical simulation	Hauet et al. (2008)
LSPIV instantaneous discharge measurements during high flows compared with rating curve & current meter reference values	Relative error at a number of observation times	**47 % at low flows, 13-23 % on rising limb, 2 % during stable high flow period**	River Arc, France, during dam release operation. Discharge range 10-150 m^3 s^{-1}, width 60-70 m, gravel-bed river.	Jodeau et al. (2008)
Microwave & UHF Doppler Radars uncertainty in instantaneous discharge measurement	Correlation coefficients between radar measurements & conventional rating curve methods over 16-week period	0.883, 0.969, 0.992 dependent on Doppler radar system	Cowlitz River, Washington, USA(5800 km^2). Width 92 m, depth 2-7 m.	Costa et al. (2006)

ADV = acoustic Doppler velocimetry; ADCP = acoustic Doppler current profiling; LSPIV = Large Scale Particle Image Velocimetry; PDF = probability density function; RMSE = root mean square error; SD = standard deviation

Table 3a: Typical quantitative results of water quality uncertainty studies: Suspended solids. Bold values were used in Figure 3.

Uncertainty Type	Estimation Method	Magnitude	Location	Reference
Instantaneous concentration	Relative difference between auto & manual dublicates	**Auto sample within 10 % of manual sample**	Devon, UK	Walling & Teed (1971)
8-year load; effect of estimation method	Bias relative to reference load from daily data (1974/75-1981/82); 12 methods tested; 6 sampling frequencies simulated via sub-sampling	**-22 to 10 %**	Euphrates (444000 km^2) at Haditha, Iraq. Ave. annual precipitation <100 mm (South) – 800 mm (north), ave. annual discharge 776 $m^3 s^{-1}$, ave. annual sediment load 1.4 10^7 t.	Al-Ansari et al. (1988); values calculated from original absolute values
8-year load; effect of sampling frequency		**-4 to 6 %**		
Instantaneous concentration; effect of cross-section sampling method	Average coefficient of variation with respect to depth- & width-integrated reference concentration	**25 %**	Various in USA	Horowitz et al. (1990); values calculated from original table
Instantaneous concentration; horizontal cross-section variation	Average coefficient of variation with respect to 5-point average	**26 %**		
Instantaneous concentration; sampler effect	Difference between two samplers (EPIC – USGS)	**36 % initially, then -1 to -15 %**	Humber catchment, UK	Evans et al. (1997); values gleaned from original graph
Concentration exceedance frequency; effect of distribution assumption given censored data	Absolute difference between lognormal & normal models	0-3%, increasing with censoring	Little Cataraqui Creek (45 km^2), Kingston Township, Ontario, Canada. Half urban, half forested, flat, ave. annual precipitation 900 mm (~22% snow).	van Buren et al. (1997)
Load; effect of distribution assumption given censored data	Relative difference between lognormal & normal models, relative to lognormal model	25-37 % (calculated from original table)		
Instantaneous load; horizontal & vertical cross-section variation	Error of point turbidity measurement compared to width- & depth- integrated sample	**-2.18 to -14.3 %**	Humber catchment, UK, 8 sites (484.3-8231 km^2). Wide range of geology, climate, soils and land cover, ave.	Wass & Leeks (1999); values from original table

			annual precipitation 600 (east) – 1600 (Pennine Hills) mm.	
5-year load; effect of rating curve choice and sampling frequency	Bias relative to reference load from daily data (1979-1983); 4 rating curves tested; 4 sampling frequencies simulated via sub-sampling	**-56 to 10 %**	Rhine catchment above Rees, Germany (165000 km^2), 5 locations. Temperate climate, 600 (lower Rhine) – 2500 (Alpes) mm precipitation, ave. annual discharge 2300 $m^3 s^{-1}$, ave. annual sediment load 3.14 10^6 t.	Asselman (2000)
Annual & 5-year load; effect of rating curve choice and sampling frequency	Bias relative to reference load from daily data; 4 rating curves tested; 12 subsets of data used to construct rating curves; various sampling frequencies simulated via sub-sampling	**WY 1996-2000: -7 to 6 % at 50 d down to -3 % at 1 d WY 1989 (low flow year): -10 to 3 % at 30 d down to -6 % at 1 d WY 1995 (median flow year): -11 to 7 % at 30 d down to -1 % at 1 d WY1982 (high flow year): -11 to 8 % at 30 d down to 3 % at 1 d**	Mississippi at Thebes, Illinois, USA (1847188 km^2), 01/10/1980-30/09/2000	Horowitz (2003); values gleaned from original graphs
		WY 1989-1993: -7 to 13 % at 50 d down to 2 % at 1 d WY 1976 (low flow year): -11 to 10 % at 50 d down to 0 % at 1 d WY 1980 (median flow year): -15 to 5 % at 30 d down to -3 % at 1 d WY1987 (high flow year): -15 to 10 % at 30 d down to -5 % at 1 d	Rhine at Maxau, Germany (50200 km^2), 31/10/1973-30/10/1993	
Annual load; effect of temporal sampling method	Relative error with respect to reference method (composite sampling)	**-9.1 to 2.7 %**	USDA-ARS Grassland Soil & Water Research Laboratory (4.6-125.1	Harmel & King (2005)

			ha), Texas, USA. Vertisol soil, 2-4 % slope, mixed land cover.	
Storm load; effect of minimum flow threshold for sampling	Professional judgement based on Harmel et al. (2002)	**±1-81 %**		Harmel et al. (2006)
Storm load; uncertainty due to manual sampling		**±15-50 % & more**		Quoted in Harmel et al. (2006): Slade (2004)
Storm load; uncertainty due to automatic sampling (intake)		**14-33 %**		Quoted in Harmel et al. (2006): Martin et al. (1992)
Storm load; uncertainty due to automatic sampling (timing)		**-65 to 51 %**		Quoted in Harmel et al. (2006)
Storm load; analytical uncertainty	95 % confidence interval	**-9.8 to 5.1 % (sandy); -5.3 to 4.4 % (fine)**		Quoted in Harmel et al. (2006): Gordon et al. (2000)
Annual load; effect of sampling frequency	Bias relative to reference load from daily data (1961-1988); 28 sampling frequencies (2-30 d) simulated via sub-sampling (50 repeats, multiple years)	**±30 % at 30 d (central 80 % from repeats & multiple years); decreasing with increasing sampling frequency**	Mississippi at St Louis, Missouri, USA (251121 km^2). Ave. annual discharge 20.1 $l\ s^{-1}\ km^{-2}$, ave. annual sediment load 447 $t\ yr^{-1}\ km^{-2}$.	Moatar et al. (2006); values gleaned from original graph; results from 35 more stations in USA and EU reported as well
Instantaneous concentration	Coefficient of variation between dublicates	**33 % (at 15 mg l^{-1}); 10 % (at 242 mg l^{-1}); 0.76 % (at 1707 mg l^{-1})**		Rode & Suhr (2007)
Analytical errors	PDF, coefficient of variation	Lognormal, 13 %		Quoted in Rode & Suhr (2007)
Storm load; effect of estimation method	Bias relative to reference load from 1-6 h data (2 events in Sep 1994 & Nov 1999); 6 estimation methods tested; continuous thinning of data down to 1 sample per event	**-52 to 19 %**	Vène catchment, France (67 km^2). Karst geology overlain by clay, mixed fruit/vegetables and urban land cover.	Salles et al. (2008); values gleaned from original graphs
Storm load; effect of sampling frequency		**-25 to 30 % at 1 sample per event; decreasing exponentially with increasing sampling frequency**		
Instantaneous concentration	Absolute difference between auto & manual dublicates	0-100 mg l^{-1}; decreasing with flow	Rowden Experimental Research Platform (1 ha fields), Devon, UK. Dystric Gleysol soil, 7-9	Krueger et al. (2009)

			% slope, grassland, ave. annual precipitation 1055 mm, 250 x 37 cm weir box.	
Storm concentrations & load	Total probable error (median in parentheses) based on RMSE propagation method	**12-26(18) % (concentrations)** **15-35(20) % (load)**	Various in USA (2.2-5506 ha)	Harmel et al. (2009) based on Harmel et al. (2006)
Concentration exceedance frequency	Uncertainty range based on bootstrapping low resolution data	Approx.10 %	Den Brook catchment (48 ha), Devon, UK. Dystric Gleysol soil, intensive grazing, ave. annual precipitation 1050 mm, flashy response, underdrained.	Bilotta et al. (2010); values gleaned from original graph
Flow-weighted mean concentration (hourly)	Trapezoidal fuzzy number based on analysis of bulk uncertainty as function of number of sub-samples for three timesteps	**±10 % core (5-6 samples per hour)** **±50 % support (1 sample per hour)**		Krueger et al. (*in press*)

Table 3b: Typical quantitative results of water quality uncertainty studies: Phosphorus. Bold values were used in Figure 3.

Uncertainty Type	Estimation Method	Magnitude	Location	Reference
Annual load; effect of sampling frequency	8 d routine sampling compared to 2 h composite (8 15 min sub-samples; Nov 1974 – May 1975); all via rating curve	**Bias -43 % (TP)**; 12 % (SRP)	River Main at Andraid, Co. Antrim, Northern Ireland (709 km^2). Basaltic glacial till geology, 10% arable, 53% grassland, 24% rough grazing, population 54549 (65% connected to sewer), ave. annual precipitation 1181 mm, flashy response.	Stevens & Smith (1978)
Annual load; effect of estimation method & sampling frequency	Bias relative to reference load from daily data (Mar 1976 to 28 Feb 1977); 3 sampling frequencies	**Average bias, biweekly: -2 to 20 %** **Average bias, bi-weekly biased to high flows: 0-2 %**	Grand River at Eastmanville, Michigan, USA (13550 km^2).	Dolan et al. (1981); values calculated from original absolute values

	simulated via sub-sampling (222-680 repeats); 3-11 estimation methods tested	**Average bias, bi-weekly biased to low flows: -1 to 2 %**	Cropland; ave. discharge 101 $m^3 s^{-1}$; ave. annual TP load 1730 kg P d^{-1}.	
Annual load (TP); effect of estimation method & sampling frequency	Bias relative to interpolated stage-triggered instantaneous load timeseries (2-15 min during rising stage, 1-4 h during falling stage, 4-24 h during baseflow); 13 estimation methods tested; 7 sampling frequencies simulated via sub-sampling	**-50 to 150 % at 12 samples per year down to -30 to 40 % at 104 samples per year; high-flow biased stratified sampling more biased and less precise**	Gelbæk catchment (8.5 km^2), Eastern Jutland, Denmark. Lowland, low baseflow, high event-responsiveness, ave. discharge 232 mm.	Kronvang & Bruhn (1996); results gleaned from original graphs
		-30 to 110 % at 12 samples per year down to -10 to 10 % at 104 samples per year; high-flow biased stratified sampling more biased and less precise	Gjern Å catchment (103 km^2), Eastern Jutland, Denmark. Lowland, high baseflow, low event-responsiveness, ave. discharge 361 mm.	
Instantaneous concentration; analytical uncertainty	Standard uncertainty (square root of variance)	0.25 µg l^{-1} (FRP(0.2 µm)) 0.32 µg l^{-1} (TP)	Latrobe River catchment, Victoria, Australia	Lovell et al. (2001)
Instantaneous concentration; spot sampling uncertainty	Standard uncertainty (square root of variance) based on 3 repeats	2.09 µg l^{-1} (FRP(0.2 µm)) 1.05 µg l^{-1} (TP)		
Instantaneous concentration; effect of spatial variation within 100 m reach	Standard uncertainty (square root of variance) based on 6 sampling spots	20.8 µg l^{-1} (FRP(0.2 µm)) 18.6 µg l^{-1} (TP)		
Annual load; effect of temporal sampling method	Relative error with respect to reference method (composite sampling)	-9.2 to 2 % (PO_4-P)	USDA-ARS Grassland Soil & Water Research Laboratory (4.6-125.1 ha), Texas, USA. Vertisol soil, 2-4 % slope, mixed land cover.	Harmel & King (2005)
Storm load; effect of minimum flow threshold for sampling	Professional judgement based on Harmel et al. (2002)	**±1-81 %**		Harmel et al. (2006)
Storm load; uncertainty due to manual sampling		±5-25 % (dissolved); ±15-50 % & more (suspended)		Quoted in Harmel et al. (2006): Slade (2004)

Storm load; uncertainty due to automatic sampling (intake)		**0-17 % (TP)**; 0 % (DP)		Quoted in Harmel et al. (2006): Martin et al. (1992)
Storm load; uncertainty due to automatic sampling (timing)		**-65 to 51 %**		Quoted in Harmel et al. (2006)
Storm load; effect of sample preservation & storage		**-64 to 92 % (TP)**; -52 to 600 % (DP)		
Storm load; analytical uncertainty		Up to ±400 % (DP); -2 to 16 % (PP)		
Flow-weighted mean concentration (TIP, weekly)	Triangular fuzzy number	±40 % support	Crighton Royal Farm (0.5 ha fields), Dumfries, Scotland, UK. Silty clay loam soil, grassland, macropore flow, ave. annual precipitation 1054 mm.	Beven et al. (2006)
Total uncertainty	PDF, mean, SD	Normal, 0, 12 % (TP)	Odense basin (1190 km^2), Denmark. Glacial/interglacial sediment geology, low rolling hills, ave. annual precipitation/evapotranspiration 900/600 mm.	Refsgaard et al. (2006)
Total analytical uncertainty	SD based on lab standards	5-15 % (PO_4-P), decreasing with concentration	2 streams in Victoria, Australia, 1 forested, 1 urbanised.	Hanafi et al. (2007)
Instantaneous concentration; horizontal cross-section variation	Coefficient of variation with respect to 10-point cross-section average	7 % (SRP)	Elbe river at Dom Muehlenholz, Germany	Rode & Suhr (2007)
Analytical errors	PDF, coefficient of variation	Normal, 6 % (TP, SRP)		Quoted in Rode & Suhr (2007): Clesceri et al. (1998)
Daily load	Total probable error based on	**<10 % (TP)**	Various in Illinois,	Gentry et al. (2007) based

	RMSE propagation method		USA. Glacial moraine geology, Mollisol soil, flat, mainly corn & soybean land cover, underdrained.	on Harmel et al. (2006)
Instantaneous concentration; analytical uncertainty	Difference to quality control standard	**±5 %**	Lough Mask catchment, Ireland	Donohue & Irvine (2008)
Instantaneous concentration; effect of lab sub-sampling	Coefficient of variation with respect to 3-sub-sample average (95 % confidence interval)	**6.4-8 % (TP)**; 6.1-7.5 % (SRP) (both almost 100 % attributable to sub-sample variability)		
Instantaneous concentration; effect of lab sub-sampling	Mean minimum detectable difference between mean concentrations of two sets of 10 replicate sub-samples from same sample	2 $\mu g\ l^{-1}$ (TP); 0.4 $\mu g\ l^{-1}$ (SRP); gleaned from original graphs		
Storm load (TP); effect of estimation method	Bias relative to reference load from 1-6 h data (2 events in Sep 1994 & Nov 1999); 6 estimation methods tested; continuous thinning of data down to 1 sample per event	**-38 to 36 %**	Vène catchment, France (67 km^2). Karst geology overlain by clay, mixed fruit/vegetables and urban land cover.	Salles et al. (2008); values gleaned from original graphs
Storm load; effect of sampling frequency		**-25 to 30 % (TP, PP)**, -25 to 65 % (SRP) at 1 sample per event; decreasing exponentially with increasing sampling frequency		
Storm concentrations & load	Total probable error (median in parentheses) based on RMSE propagation method	13-103(19) % (PO_4-P concentrations); 14-104(23) % (PO_4-P load); **16-104(24) % (TP concentrations); 17-105(27) % (TP load)**	Various in USA (2.2-5506 ha)	Harmel et al. (2009) based on Harmel et al. (2006)
Concentrations & load	Total probable error based on RMSE propagation method	27 % (PO_4-P concentrations); 28 % (PO_4-P load)		Quoted in Harmel et al. (2009): Keener et al. (2007)
Instantaneous concentration (TP)	Absolute difference between auto & manual dublicates	0-400 $\mu g\ l^{-1}$; decreasing with flow	Rowden Experimental Research Platform (1 ha fields), Devon, UK. Dystric Gleysol soil, 7-9 % slope, grassland, ave.	Krueger et al. (2009)

			annual precipitation 1055 mm, surface soil P ~540 mg kg^{-1}, 250 x 37 cm weir box.	
Annual load; effect of sampling frequency	Bias relative to reference load from stratified data (2-4 per d when dry, up to 8 per d when wet; Feb 2005 – Jan 2006); 5 sampling frequencies simulated via sub-sampling	**Monthly: -21.3 to 35.2 % (TP)**; -10.6 to 27.9 % (SRP) **Fortnightly: -17.5 to 28.1 % (TP)**; -11 to 15.3 % (SRP) **Weekly: -11.6 to 15.4 % (TP)**; -4.9 to 6.5 % (SRP) **Daily: 0 to 4 % (TP)**; -2.1 to 2.5 % (SRP) **12h: -1.9 to 0.7 % (TP)**; -0.9 to 1.1 % (SRP)	Frome at East Stoke, UK (414 km^2), Mainly chalk geology, mainly grassland & cereals land cover, one town, ave. annual precipitation 1020 mm, ave. annual discharge 6.38 m^3 s^{-1}, BFI 0.84.	Bowes et al. (2009)
Precision of various high frequency nutrient analysers	As stated by manufacturer	±2 % of range (PO_4-P, Greenspan™ Aqualab; PO_4-P, Ecotech™ FIA NUT1000; PO_4-P, FIALab™ SIA) **±3 % of range (TP & PO_4-P, Systea™ Micromac C; PO_4-P, Envirotech™ AutoLAB/MicroLAB)**		Bende-Michl & Hairsine (2010)
Annual load (TP); effect of temporal sampling method	Bias relative to interpolated 20 min instantaneous load timeseries	**Median bias of various methods -50 to +30 %**	Co. Monaghan, Ireland (5 km^2). Drumlin soils, grassland, flashy, point sources.	Jordan & Cassidy (2011)
Flow-weighted mean concentration (TP, hourly)	Trapezoidal fuzzy number based on analysis of bulk uncertainty as function of number of sub-samples for three timesteps	**±10 % core (5-6 samples per hour)** **±50 % support (1 sample per hour)**	Den Brook catchment (48 ha), Devon, UK. Dystric Gleysol soil, intensive grazing, ave. annual precipitation 1050 mm, flashy response, underdrained.	Krueger et al. (*in press*)

Table 3c: Typical quantitative results of water quality uncertainty studies: Nitrogen. Bold values were used in Figure 3.

Uncertainty Type	Estimation Method	Magnitude	Location	Reference
Annual load (NO_3-N); effect of sampling frequency	8 d routine sampling compared to 2 h composite (8 15 min sub-samples; Nov 1974 – May 1975); all via rating curve	Bias 18 %	River Main at Andraid, Co. Antrim, Northern Ireland (709 km^2). Basaltic glacial till geology, 10% arable, 53% grassland, 24% rough grazing, population 54549 (65% connected to sewer), ave. annual precipitation 1181 mm, flashy response.	Stevens & Smith (1978)
Annual load (TN); effect of estimation method & sampling frequency	Bias relative to interpolated stage-triggered instantaneous load timeseries (2-15 min during rising stage, 1-4 h during falling stage, 4-24 h during baseflow); 13 estimation methods tested; 7 sampling frequencies simulated via sub-sampling	**-20 to 30 % at 12 samples per year down to -12 to 10 % at 104 samples per year; high-flow biased stratified sampling more biased and less precise**	Gelbæk catchment (8.5 km^2), Eastern Jutland, Denmark. Lowland, low baseflow, high event-responsiveness, ave. discharge 232 mm.	Kronvang & Bruhn (1996); results gleaned from original graphs
		-11 to 25 % at 12 samples per year down to -2 to 9 % at 104 samples per year; high-flow biased stratified sampling more biased and less precise	Gjern Å catchment (103 km^2), Eastern Jutland, Denmark. Lowland, high baseflow, low event-responsiveness, ave. discharge 361 mm.	
Annual load (NO_3-N); effect of temporal sampling method	Relative error with respect to reference method (composite sampling)	-9.2 to 2 %	USDA-ARS Grassland Soil & Water Research Laboratory (4.6-125.1 ha), Texas, USA. Vertisol soil, 2-4 %	Harmel & King (2005)

			slope, mixed land cover.	
Storm load; effect of minimum flow threshold for sampling	Professional judgement based on Harmel et al. (2002)	**±1-81 %**		Harmel et al. (2006)
Storm load; uncertainty due to manual sampling		±5-25 % (dissolved); ±15-50 % & more (suspended)		Quoted in Harmel et al. (2006): Slade (2004)
Storm load; uncertainty due to automatic sampling (intake)		**0 % (TN)**; 0-4 % (DN)		Quoted in Harmel et al. (2006): Martin et al. (1992)
Storm load; uncertainty due to automatic sampling (timing)		**-65 to 51 %**		Quoted in Harmel et al. (2006)
Storm load; effect of sample preservation & storage		-90 to 83 % (NH_3-N); -65 to 71 % (NO_3-N); -84 to 49 % (TKN)		
Storm load; analytical uncertainty		Up to ±400 % (DN); ±4-30 % (PN)		
Total uncertainty (TN)	PDF, mean, SD	Normal, 0, 10 %	Odense basin (1190 km^2), Denmark. Glacial/interglacial sediment geology, low rolling hills, ave. annual precipitation/evapotranspiration 900/600 mm.	Refsgaard et al. (2006)
Total analytical uncertainty (NH_4-N)	SD based on lab standards	4-19 %, decreasing with concentration	2 streams in Victoria, Australia, 1 forested, 1 urbanised.	Hanafi et al. (2007)
Instantaneous concentration (NO_3-N); analytical uncertainty	SD	0, 40, 50, 50 μg l^{-1} at 100, 200, 800, 2100 μg l^{-1}, respectively		Rode & Suhr (2007)
Instantaneous concentration (NH_4-N); analytical uncertainty	Mean SD	5-8 %		
Instantaneous concentration (NH_4-N); horizontal cross-section variation	Variation from 10-point cross-section average	Up to 50 % & more	Elbe river at Dom Muehlenholz, Germany	

Analytical errors	PDF, coefficient of variation	Normal, 5 % (NO_3, Cadmium Reduction Method); normal, 2.5 % (NO_3, Electrode Method); normal, 4 % (NO_3, Ion Chromatography); normal, 6 % (NO_2); normal, 11 % (NH_4)		Quoted in Rode & Suhr (2007): Clesceri et al. (1998)
Instantaneous concentration; analytical uncertainty	Difference to quality control standard	**±5 %**	Lough Mask catchment, Ireland	Donohue & Irvine (2008)
Instantaneous concentration; effect of lab sub-sampling	Coefficient of variation with respect to 3-sub-sample average (95 % confidence interval)	**9.6-11.2 % (TN)**, 71.8-82 % (lakes) & 77-82.2 % (rivers) attributable to sub-sample variability; 4-6.6 % (DIN), 53.4-71.2 % (lakes) & 67.7-75.1 % (rivers) attributable to sub-sample variability		
Instantaneous concentration; effect of lab sub-sampling	Mean minimum detectable difference between mean concentrations of two sets of 10 replicate sub-samples from same sample	0.2 mg l^{-1} (TN); 0.02 mg l^{-1} (DIN); gleaned from original graphs		
Storm load (TN); effect of estimation method	Bias relative to reference load from 1-6 h data (2 events in Sep 1994 & Nov 1999); 6 estimation methods tested; continuous thinning of data down to 1 sample per event	**-22 to 11 %**	Vène catchment, France (67 km^2). Karst geology overlain by clay, mixed fruit/vegetables and urban land cover.	Salles et al. (2008); values gleaned from original graphs
Storm load; effect of sampling frequency		**-25 to 20 % (TN)**, -25 to 10 % (NO_3-N) at 1 sample per event; decreasing exponentially with increasing sampling frequency		
Storm concentrations & load	Total probable error (median in parentheses) based on RMSE propagation method	13-102(17) % (NO_3-N concentrations); 14-103(22) % (NO_3-N load); **14-104(23) % (TN concentrations); 15-105(27) % (TN load)**	Various in USA (2.2-5506 ha)	Harmel et al. (2009) based on Harmel et al. (2006)
Annual load (TON); effect of sampling frequency	Bias relative to reference load from stratified data (2-4 per d when dry, up to 8 per d when wet; Feb 2005 – Jan 2006); 5 sampling frequencies simulated via sub-sampling	-4.2 to 11.2 % (monthly); -3.5 to 3.9 % (fortnightly); -1.8 to 3.5 % (weekly); -0.5 to 0.9 % (daily); -0.1 to 0.3 % (12 h)	Frome at East Stoke, UK (414 km^2), Mainly chalk geology, mainly grassland & cereals land cover, one town, ave.	Bowes et al. (2009)

			annual precipitation 1020 mm, ave. annual discharge 6.38 $m^3 s^{-1}$, BFI 0.84.	
Precision of various high frequency nutrient analysers	As stated by manufacturer	±5 % of range (NH_4-N & NO_3-N, WTW™ VARiON; NH_4-N & NO_3-N, Greenspan™ Aqualab; NO_3-N, YSI™ YSI96000) **±3 % of range (TN, NH_4-N, NO_3-N & NO_2-N, Systea™ Micromac C; NO_3-N & NO_2-N, S::can™ Spectroanalyser)** ±2 % of range (NH_4-N & NO_3-N, Envirotech™ AutoLAB/MicroLAB; NH_4-N, NO_3-N & NO_2-N, FIALab™ SIA; NO_3-N, Satlantic™ ISUS)		Bende-Michl & Hairsine (2010)

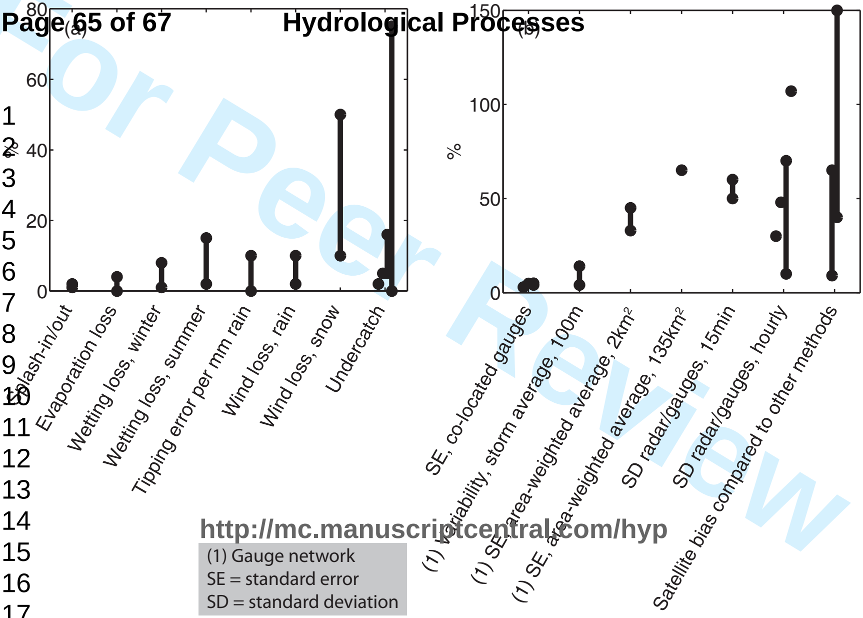
BFI = base flow index; DIN = dissolved inorganic nitrogen; DN = dissolved nitrogen; DP = dissolved phosphorus; FRP(X µm) = filtered reactive phosphorus (filter size); PDF = probability density function; PN = particulate nitrogen; PP = particulate phosphorus; RMSE = root mean square error; SD = standard deviation; SRP = soluble reactive phosphorus; TIP = total inorganic phosphorus; TKN = total Kjeldahl nitrogen; TN = total nitrogen; TP = total phosphorus; WY = water year

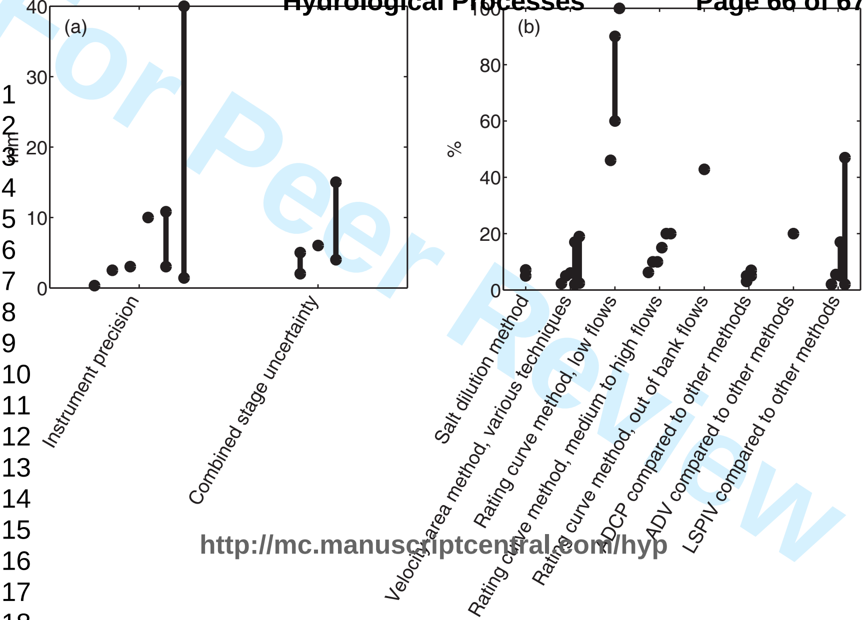
FIGURE CAPTIONS

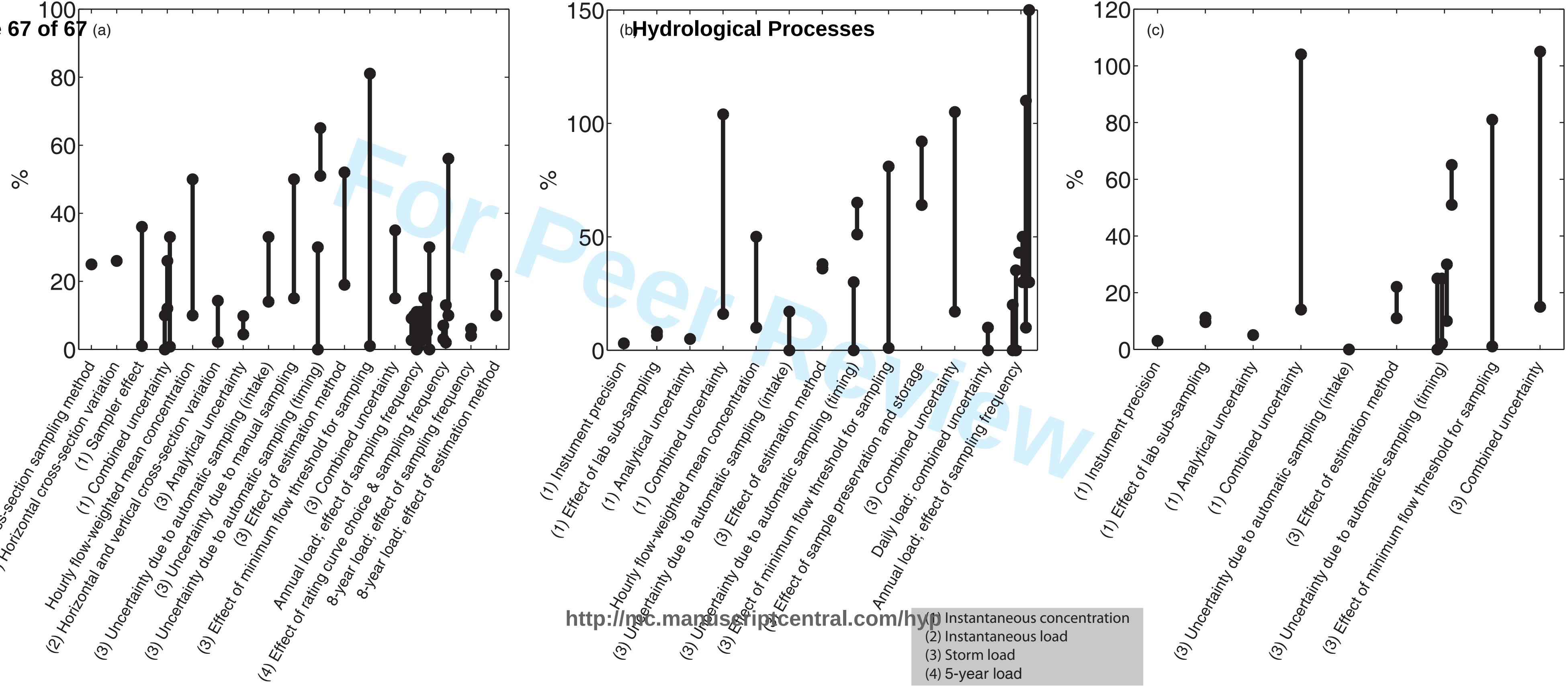
Figure 1: (a) Typical ranges of raingauge error components. (b) Typical ranges of combined rainfall uncertainty across spatial scales (from gauge to satellite footprint). See Table 1a-1c for details.

Figure 2: (a) Typical ranges of stage measurement uncertainty. (b) Typical ranges of combined discharge uncertainty from various methods. See Table 2a-2c for details.

Figure 3: Typical uncertainty ranges across temporal scales (from instantaneous concentration to multi-annual load) for: (a) suspended solids, (b) total phosphorus, (c) total nitrogen. See Table 3a-3c for details.


(a)
80
60
40
20
0
%
Splash-in/out
Evaporation loss
Wetting loss, winter
Wetting loss, summer
Tipping error per mm rain
Wind loss, rain
Wind loss, snow
Undercatch
(b)
150
100
50
0
%
SE, co-located gauges
(1) variability, storm average, 100m
(1) SE, area-weighted average, 2km²
(1) SE, area-weighted average, 135km²
SD radar/gauges, 15min
SD radar/gauges, hourly
Satellite bias compared to other methods
(1) Gauge network
SE = standard error
SD = standard deviation


(a)
mm
0
10
20
30
40
Instrument precision
Combined stage uncertainty
(b)
%
0
20
40
60
80
100
Salt dilution method
Velocity area method, various techniques
Rating curve method, low flows
Rating curve method, medium to high flows
Rating curve method, out of bank flows
ADCP compared to other methods
ADV compared to other methods
LSPIV compared to other methods


(a)
%
100
80
60
40
20
0
(1) Effect of cross-section sampling method
(1) Horizontal cross-section variation
(1) Sampler effect
(1) Combined uncertainty
Hourly flow-weighted mean concentration
(2) Horizontal and vertical cross-section variation
(3) Analytical uncertainty
(3) Uncertainty due to automatic sampling (intake)
(3) Uncertainty due to manual sampling
(3) Uncertainty due to automatic sampling (timing)
(3) Effect of estimation method
(3) Effect of minimum flow threshold for sampling
(3) Combined uncertainty
Annual load; effect of sampling frequency
(4) Effect of rating curve choice & sampling frequency
8-year load; effect of sampling frequency
8-year load; effect of estimation method
(b)
%
150
100
50
0
(1) Instument precision
(1) Effect of lab sub-sampling
(1) Analytical uncertainty
(1) Combined uncertainty
Hourly flow-weighted mean concentration
(3) Uncertainty due to automatic sampling (intake)
(3) Effect of estimation method
(3) Uncertainty due to automatic sampling (timing)
(3) Effect of minimum flow threshold for sampling
(3) Effect of sample preservation and storage
(3) Combined uncertainty
Daily load; combined uncertainty
Annual load; effect of sampling frequency
(c)
%
120
100
80
60
40
20
0
(1) Instument precision
(1) Effect of lab sub-sampling
(1) Analytical uncertainty
(1) Combined uncertainty
(3) Uncertainty due to automatic sampling (intake)
(3) Effect of estimation method
(3) Uncertainty due to automatic sampling (timing)
(3) Effect of minimum flow threshold for sampling
(3) Combined uncertainty
(1) Instantaneous concentration
(2) Instantaneous load
(3) Storm load
(4) 5-year load