



Large-sample hydrology: a need to balance depth with breadth

H. V. Gupta¹, C. Perrin², G. Blöschl³, A. Montanari⁴, R. Kumar⁵, M. Clark⁶, and V. Andréassian²

¹Department of Hydrology and Water Resources, The University of Arizona, Tucson, AZ, USA

²Irstea, Hydrosystems and bioprocesses Research Unit (HBAN), Antony, France

³Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria

⁴Department DICAM, University of Bologna, Bologna, Italy

⁵UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany

⁶Hydrometeorological Applications Program, Research Applications Laboratory, Boulder, CO, USA

Correspondence to: H. V. Gupta (hoshin.gupta@hwr.arizona.edu)

Received: 26 June 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 12 July 2013

Revised: 18 December 2013 – Accepted: 26 December 2013 – Published: 6 February 2014

Abstract. A holy grail of hydrology is to understand catchment processes well enough that models can provide detailed simulations across a variety of hydrologic settings at multiple spatiotemporal scales, and under changing environmental conditions. Clearly, this cannot be achieved only through intensive place-based investigation at a small number of heavily instrumented catchments, or by empirical methods that do not fully exploit our understanding of hydrology. In this opinion paper, we discuss the need to actively promote and pursue the use of a “large catchment sample” approach to modeling the rainfall–runoff process, thereby balancing depth with breadth. We examine the history of such investigations, discuss the benefits (improved process understanding resulting in robustness of prediction at ungauged locations and under change), examine some practical challenges to implementation and, finally, provide perspectives on issues that need to be taken into account as we move forward. Ultimately, our objective is to provoke further discussion and participation, and to promote a potentially important theme for the upcoming Scientific Decade of the International Association of Hydrological Sciences entitled *Panta Rhei*.

1 Introduction

“Because almost any model with sufficient free parameters can yield good results when applied to a short sample from a single catchment, effective testing requires that models be tried on many catchments of widely differing characteristics,

and that each trial cover a period of many years.” (Linsley, 1982).

1.1 Motivations for developing large-sample hydrology

A holy grail of hydrological science is to achieve a degree of process understanding that enables construction of models that are capable of providing detailed and physically realistic simulations across a variety of different hydrologic environments, and at multiple spatial and temporal scales (Nash and Sutcliffe, 1970; Klemeš, 1986a; Michel et al., 2006). With its focus on reducing predictive uncertainty, the Prediction in Ungauged Basins (PUB) initiative of the International Association of Hydrological Sciences (IAHS) has helped move the culture of hydrologic science closer to this objective, and away from a reliance upon universal models applied via recalibration at each study location (Hrachowitz et al., 2013).

This move has deep roots (see Linsley, 1982), but has become considerably stronger since the 1999 IAHS meeting in Birmingham where the idea was extensively discussed. It has helped drive the search for improved understanding of the hydrological cycle, and for modeling approaches that

- achieve the three R’s (reliability, robustness and realism);
- have greater generality and transposability; and for which
- the parameters can be more easily specified from data.

Clearly, this cannot be achieved *only* through detailed studies at heavily instrumented catchments – although such studies are of critical importance. Nor can it be achieved by simple regionalization methods based primarily in statistical approaches and/or spatial extrapolation rather than improved understanding of hydrologic behavior. What is needed is to take further advantage of the extensive data sets now available (and becoming available) to make common a large-sample approach to hydrological investigation (Andréassian et al., 2006). While this emphasis has long been the focus of what is known as comparative hydrology (e.g., Kovacs, 1984; Falkenmark and Chapman, 1989; Thompson et al., 2011; Blöschl et al., 2013; see also Sect. 5.2), it is gratifying to see that more and more recent investigations have recognized the value of adopting this approach (see numerous examples cited in this paper).

1.2 The context of current practice

The context of much current hydrological practice is a focus on *depth* rather than *breadth* (with several notable exceptions), wherein detailed process investigation and model development/refinement are conducted *at only one or a limited number of catchments*. The typical goal is to (a) learn more about a specific catchment by improving upon some prior concept, or (b) establish a basis for prediction and decision-making at that specific location. This might be called place-based learning. By contrast, the scientific aspiration is to generalize from the study of specific cases, so that we can discover and establish general hydrological principles, thereby advancing hydrological understanding.

Over the past several decades, the hydrology community has developed a number of generic model codes or model development frameworks (multi-hypothesis environments), that can be applied to a given catchment study (see discussion in Clark et al., 2011 and Gupta et al., 2012). These codes make it possible to select a specific off-the-shelf model structure, after which one need only specify values for the parameters. If model performance is inadequate, attempts can be made to diagnose model deficiencies (Fig. 1) and find ways to correct/improve the model/hypothesis (Gupta et al., 2008; McMillan et al., 2011; Euser et al., 2013). The community has also made available a large number of data sets that can be used to test such codes.

However, attempts at such generalization are fraught with difficulties. Faced with the tremendous geo–eco–hydro-climatic variability of environments across the world, attempts based on generic models are typically complicated by considerable noise in the individual results (e.g., Oudin et al., 2010; Savenije, 2009) – these being due to unresolved model identification issues arising from a combination of inadequate data (insufficient information), data noise, model structural inadequacy, and weak model identification techniques (e.g., see Gupta et al., 1998, 2008, 2012). This is

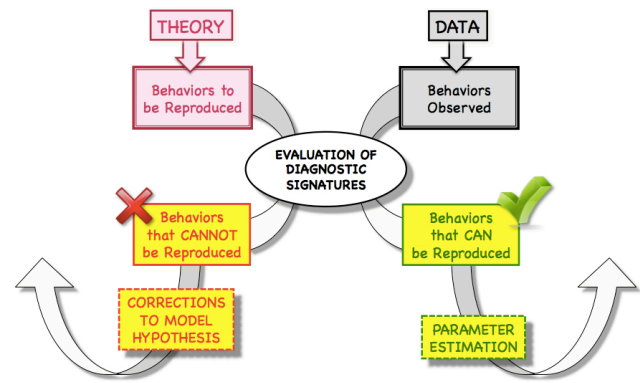


Fig. 1. How an approach based in evaluation of signature properties can be used to detect and diagnose model deficiencies and develop appropriate ways to improve the model/hypothesis (figure based on ideas presented in Gupta et al., 2008).

particularly true as both model realism and spatiotemporal resolution are increased.

On the one hand, such difficulties have contributed to the counter notion of uniqueness of place (Beven, 2000), and that the model structure should adapt to reflect spatial differences in the dominant hydrologic processes. On the other hand, global-scale hydrological studies typically use only a *single conceptual structure* to represent *all* locations around the world, while attempting to account for the place-to-place differences entirely through the specification of model parameters (representing differences in soil and vegetation properties) while largely ignoring the spatial variability in dominant hydrological processes (Dirmeyer et al., 2006).

1.3 Purpose of the paper

The purpose of this opinion paper is to encourage a greater focus on large catchment sample type studies in hydrology, to complement the approach of intensive place-based investigation. Since the use of large samples in hydrology is certainly not new, we provide some historical perspective (but do not attempt a comprehensive review), motivating the need for more such studies, illuminating some of the challenges, and examining issues related to their design and implementation. To be clear, we do not suggest a reduction in efforts dealing with detailed catchment studies; both kinds of investigations (including ones in the middle ground) are necessary. Our objective is to provoke further discussion and participation, and to promote an important theme for the upcoming IAHS scientific decade (Panta Rhei; Montanari et al., 2013).

2 Previous studies: what have we learned?

2.1 Early attempts at large-sample studies

The issue of using large numbers of catchments for hydrological investigation is not new; early attempts go back more than thirty years. However, practical reasons – including limitations in data access, computing requirements or the ability to collaborate efficiently – made such efforts difficult. Further, there was a common belief that models developed in one location could not be readily applied elsewhere. Consequently, it was difficult to really know the respective merits and usefulness of any existing model.

In 1967, the World Meteorological Organization (WMO) launched an initiative to develop an inventory of models, along with advice to users regarding their accuracy under various hydro-climatic conditions. Not surprisingly, it was quickly deemed useful to carry out an actual model inter-comparison study and, in 1973, ten simulation models from seven countries were applied to a set of six catchments (from the USA, the former Soviet Union, Australia, Japan, Cameroon and Thailand) that represent a variety of hydro-climatic conditions (WMO, 1975; Sittner, 1976). Although the investigation did not arrive at definitive conclusions regarding the merits of the models tested, it drew attention to the need for a deeper and wider evaluation of models. At that time, a small number of other model inter-comparison studies involving more than one catchment were also carried out; for example see Mein and Brown (1975) who compared three models on four catchments in Australia, and James (1972), Egbuniwe and Todd (1976) and Magette et al. (1976) for their work on the Stanford model using 2 to 16 catchments. From the 1970s onward, a significant number of studies have used large samples for the statistical investigation of floods (e.g., Hardison, 1974; Hosking et al., 1985; Lichty et al., 1990).

A few years later Linsley (1982) listed generality as one of four main properties desirable in hydrological models (along with accuracy, applicability and ease-of-use). Linsley argued that it was necessary to break out of the habitual practice of developing a different model for each catchment, because that eliminates the opportunity for learning what comes with repeated applications of the same model. He suggested the necessity for extensive testing of new models (see introductory quote) so that ones that are not sufficiently general can be eliminated, and stressed the usefulness of large-scale applications, saying the application of a model to many catchments results in many sets of parameters which can conceivably serve as a basis for objective determination of parameters from physical characteristics of the catchments.

Along the same line of thought, Klemeš (1986a) proposed a formal four-level testing scheme to evaluate transposability of a model in time and space. Despite the demanding nature of such testing, Klemeš (1986a) regarded this scheme as a minimum requirement, and stated that the “use of more test basins [...] would increase the credibility standing of

a model, and an accumulation of test results may lead to meaningful generalizations”. Bergström (1991) agreed, stating “growing confidence in hydrological modeling can be obtained by applying the model under a span of different geographical, climatological and geological conditions.”

These ideas, expressed by leading hydrologists, set the basis for studies involving large catchment samples, that have now begun to be more common.

2.2 Examples of relevant literature from a rainfall–runoff modeling perspective

While several modeling studies during the 1980s used more than one catchment (e.g., Weeks and Hebbert, 1980; Naef, 1981; Pirt and Bramley, 1985; Loague and Freeze, 1985; Weeks and Ashkanasy, 1985; WMO, 1986; Srikanthan and Goodspeed, 1988), the actual emergence of large-sample studies arguably occurred in the early 1990s, coinciding with a progressive increase in availability of computing power, and using time series that are sufficiently long to enable robust assessments.

To illustrate the growing interest in large-sample studies, a list of 94 published rainfall–runoff modeling studies that used more than 30 catchments is presented in the Supplement. This list is *not* intended to be a comprehensive or exhaustive list, but instead a rough overview, given that there are many other kinds of hydrological investigations (such as flood studies, hydro-ecological investigations, etc.) that are based in a large-sample approach, but these are not listed here. The sample sizes in our list range from 30 to 1508, with a median of 140. Early studies were mainly from Australia, France and Belgium, followed later by the UK, Austria and USA, etc. (see e.g., the reviews by Boughton, 2006 and Merz et al., 2006, for Australia and Austria respectively, and more comprehensive discussion in Blöschl et al., 2013).

The catchments in this list are from a variety of physical, climatic and hydrological conditions (see Fig. 2). While some studies focused on national data sets, others included catchments from several countries (although typically less than five). The studies focused on a range of spatial and temporal scales: catchment areas ranged from 1 to 130 000 km² and models were run at hourly, daily, monthly, annual and inter-annual time steps. The study goals included a variety of purposes, most commonly being related to

- a. model development, application and comparison;
- b. estimation of model parameters (calibration and regionalization techniques);
- c. evaluation procedures and criteria;
- d. sensitivity and uncertainty analysis; and
- e. impact assessment studies.

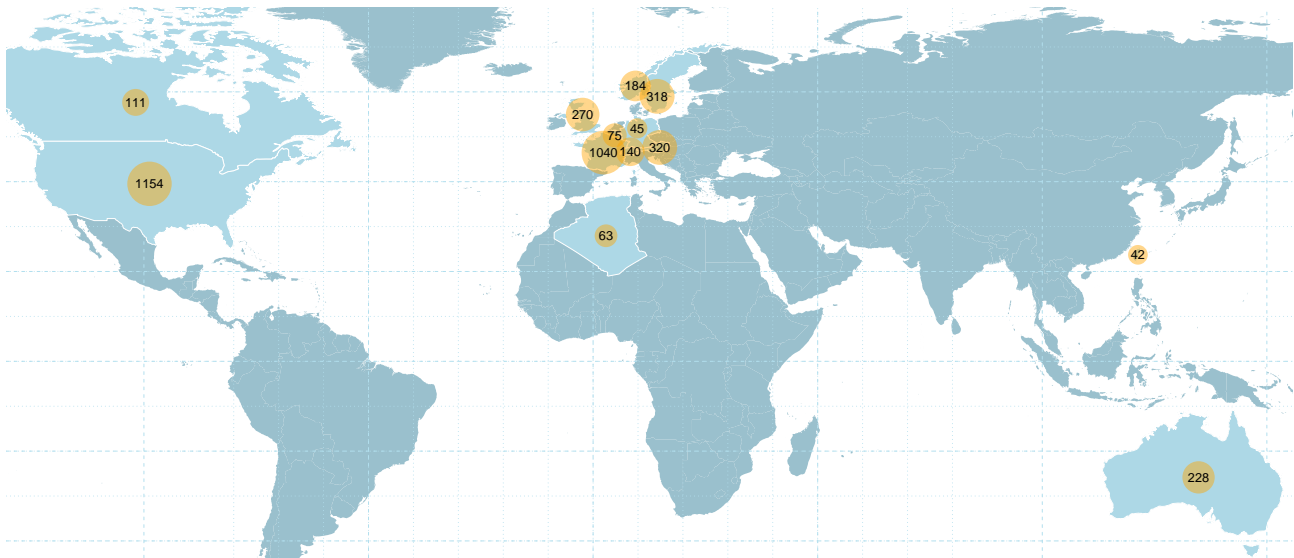


Fig. 2. Map showing locations of catchments reported in the Supplement. Size of symbol is proportional to numbers of catchments from the corresponding country.

Not surprisingly, most studies used conceptual-type rainfall–runoff (CRR) models rather than physically based models, probably due to their relative ease of implementation on large samples and lower data and computing requirements. In general, the reasons for using large catchment samples were

1. to arrive at conclusions that were more general than could be achieved using a single catchment (e.g., about the relative merits of various methods);
2. to establish range of applicability, or expected level of efficiency, of methods/models; or
3. to ensure sufficient information to enable statistically significant relationships to be established (e.g., between catchment descriptors and model parameters in regionalization studies).

In addition, several groups compiled large-sample data sets with the goal of facilitating collective efforts. The pioneering inter-comparison studies sponsored by the WMO (1975, 1986, 1992) involved several teams and data sets. The Model Parameter Experiment promoted by NOAA (MOPEX, www.nws.noaa.gov/oh/mopex/; see Schaake et al., 2001; Chahinian et al., 2006; Duan et al., 2006; Schaake and Duan, 2006, etc.) engaged several groups in the study of a sample of 438 catchments from the USA and 40 in France. Most recently, the Distributed Model Intercomparison Projects (DMIP-I and DMIP-II, Smith et al., 2004, 2012; www.nws.noaa.gov/oh/hrl/dmip/) engaged several groups in a comparative assessment of spatially distributed hydrological models (and associated parameter estimation strategies), using comprehensive data sets from a large number of US catchments.

Likewise large-scale comparative studies such as the North American Land Data Assimilation System (NLDAS-I and NLDAS-II), the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS), the Rhone-Aggregation Land Surface Scheme Intercomparison Project, and the Global Energy and Water Cycle Exchanges Project (GEWEX), have compiled a large array of hydro-meteorological data sets (see Table 1). While these data sets were compiled for regional-scale land-surface models at a relatively large spatial resolution ($1/8^\circ$ or larger), such data sets could provide a useful starting point for hydrological investigations over large and diverse spatial domains.

3 General benefits of large-sample hydrology

There are at least four clear benefits to studies that work with data from large numbers of catchments:

3.1 Improved understanding

First, large-sample studies provide better opportunities for improving hydrological science (Ehret et al., 2013) by facilitating rigorous testing of competing model structures and component hypothesis (Clark et al., 2011), and enabling better diagnosis of limitations, range of applicability and capabilities for extrapolation (Gupta et al., 2008; Martinez and Gupta, 2011; Coron et al., 2012). Recent applications of the comparative hydrology approach are an excellent example (Blöschl et al., 2013).

Table 1. Examples of data sets potentially useful for large-sample hydrological investigations.

Name	Data type	Ref/provider	URL	Availability
MOPEX	438 US catchments	NOAA	www.nws.noaa.gov/oh/mopex/	Freely available
DMIP	US catchments	NOAA	www.nws.noaa.gov/oh/hrl/dmip/	Freely available
None	60 UK catchments	PUB Top-Down Model Working Group	tdwg.catchment.org/	Freely available
None	278 Australian catchments	Peel et al. (2000)		Available upon request
GLACE-I & II	Global	Global Land–Atmosphere Coupling Experiment	gmao.gsfc.nasa.gov/research/GLACE-2/	Freely available
NLDAS-I & II	CONUS domain	NASA Goddard Earth Sciences Data and Information Services Center	http://ldas.gsfc.nasa.gov/index.php	Freely available
Rhone-AGG	Rhone basin (France)	French research community	www.cnr.meteo.fr/isbadoc/projects/rhoneagg/	Available upon request
GEWEX	Regional to Global data sets relevant to water and energy budgets	GEWEX	www.gewex.org/	Freely available
GRDC	Global runoff data	Global Runoff Data Centre	http://www.bafg.de/GRDC/EN/Home/homepagenode.html	Available with a nominal cost
USGS	Hydrology relevant US data sets (e.g., HCDN/GAUGES data sets)	USGS	www.usgs.gov/	Freely available
E-OBS	Pan-EU temperature and precipitation	European Climate Assessment & Dataset (ECA & D)	http://eca.knmi.nl/	Freely available
ORNL DAAC	Several useful regional to global data sets relevant to water and energy budgets	Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC)	http://daac.ornl.gov/index.shtml	Mostly freely available
None	Global-scale atmospheric data	Van der Ent and Savenije (2011)		Contact authors
None	Regional flux tower data	Williams et al. (2012)		Contact authors

3.2 Robustness of generalizations

Second is the possibility of bringing methods of statistical analysis to bear, so that statistical robustness can be achieved, and degrees of confidence can be established (Mathevet et al., 2006). In doing so, large samples can help to both reduce the impact of severe data errors (sometimes present in only a few

of the catchments), and to identify and target outliers (unusual cases) for special attention (Andréassian et al., 2010). In this context, it is important to not reject a data set just because the selected model fails to reproduce its behavior (Le Moine et al., 2007; Boldetti et al., 2010).

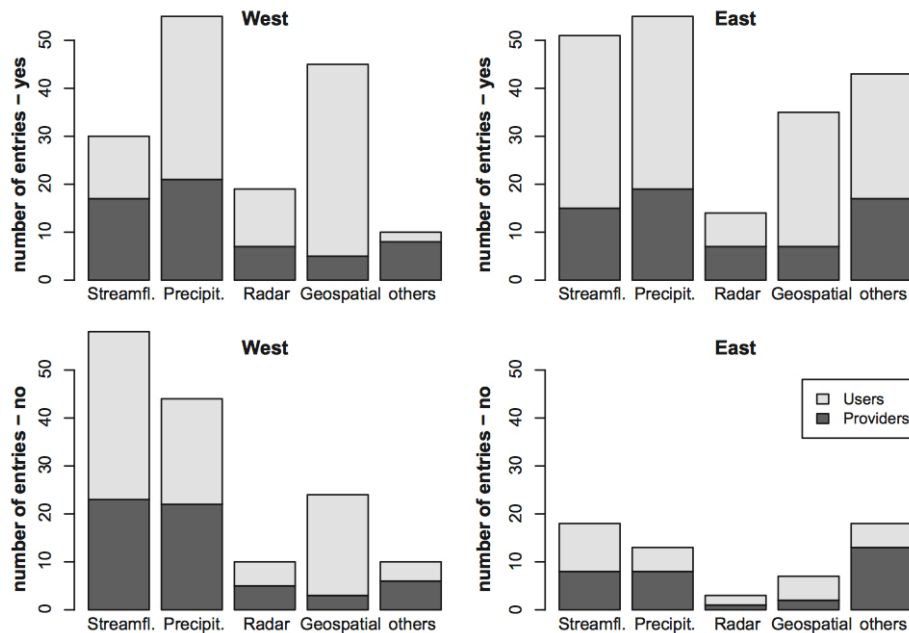


Fig. 3. Shows perceptions regarding economic barriers to the availability of hydrometeorological data in Europe. Survey results are stratified by data providers (dark grey) and data users (light grey); western Europe (left subplots) and eastern Europe (right subplots); and type of data (bars indicating streamflow, precipitation, radar, geospatial, and others). Yes responses indicate that economic barriers are perceived to exist, whereas No indicates that economic barriers are perceived not to exist (reproduced from Viglione et al., 2010).

3.3 Classification, regionalization and model transfer

Third, large samples support and facilitate the development of catchment classification and regionalization systems that provide insights into hydrological behavior (a key motivation of the PUB initiative), thereby facilitating transfer of understanding to ungauged locations (Sivapalan, 2003; Göttinger and Bardossy, 2007; Oudin et al., 2010). Since, process dominance varies with climatic and physiographical characteristics of a catchment (terrain, soil, geology, etc.) and other factors, a satisfactory regionalization system (see important work by Winter, 2001; McDonnell and Woods, 2004; Wagener et al., 2007; among others) cannot be achieved without spatially extensive data sets that are representative of catchment types worldwide. Further, prior testing on large and diverse catchment sets paves the way for improved model transfer into operational use; as expressed by Bergström (1991) “the large number of applications have gradually built up our confidence in the use of these models to a degree where we can continue our operational applications and accept the models as the foundation for further model development.”

3.4 Estimation of uncertainty

Fourth, large samples support and facilitate better understanding of how much uncertainty can be expected in model predictions given available knowledge (Andréassian et al., 2007), by making possible a statistical regionalization of

uncertainty estimates (Skøien and Blöschl, 2007; Bourgin et al., 2013; Blöschl et al., 2013) and thereby indicating (a priori) how much prediction uncertainty can be expected at arbitrary locations (including ungauged ones).

4 Challenges to practical implementation

4.1 Availability and quality of data sets having large numbers of catchments

Widespread availability of large-sample data sets is, of course, a key requirement for further progress. However, attempts to make (currently local) data sets available to the global community have run into problems related to economics and ownership. While hydrology data in the USA are mainly in the public domain and easily accessible to scientists worldwide, legal and economic barriers to exchange of data inhibit the wider spread of many national data sets (Fig. 3). While discharge data are increasingly more freely available, climate data are often less easily accessible in many countries. To better understand such barriers, Viglione et al. (2010) surveyed data providers and users from 32 European countries. The primary reason cited was economic, with public institutions responsible for data collection and administration facing growing financial pressure due to reductions in government funding (Freebairn and Zillman, 2002). Other reasons cited were conflicts of interest (as when the provider

sells products derived from the data) and the desire to reduce misuse of data due to inappropriate redistribution by users.

Nonetheless, several large catchment sample data sets are freely available. We list several examples in Table 1 that could be useful for large-sample hydrological investigations. However, there is the important issue of adequate metadata (details regarding how the data were collected, and what can be done with them). While it is increasingly becoming standard practice to provide metadata, at least for large data sets (e.g., see the WMO/GRDC initiative www.bafg.de/GRDC and wis.wmo.int, and the work on metadata for NEXRAD by Kruger et al., 2011), such information is not generally available for catchment data sets. One reason is that it is generally more difficult to apply the same kind of rigorous quality checks that can be implemented at heavily instrumented catchments, and even something as basic as visual inspection can be too time-consuming when compiling samples of hundreds of catchments.

Finally, as pointed out by a reviewer we should mention that there is a considerable amount of relevant grey literature dealing with catchment data sets from applied sciences and operational practice (e.g., NOAA, 2007; Hughes, 2013; Uhlemann et al., 2013). We face a significant challenge in better exploiting and incorporating such data into scientific investigation.

4.2 Reporting and sharing protocols for data and models

A longstanding issue is, of course, the need for coherence in the way data and models are reported, stored and shared; this is currently done in a number of different ways depending on their nature and the purpose for which they were compiled. Data deemed to be of wider interest to the hydrologic community are now increasingly published in journals through data and analysis notes, along with metadata. Further, data collected by Hydro-meteorological Services, and other public agencies, are more easily accessible via the Internet, although typically with much less meta-information (Viglione et al., 2010).

While considerable attention has been given to protocols for documenting and sharing data during the past several decades (Jones et al., 1979; Goodall et al., 2008; Viglione et al., 2010; see also <https://www.wmo.int/pages/prog/gcos/Publications/gcos-96.pdf>), the procedures for documenting and sharing models (computer codes) continue to remain extremely ad hoc. At the same time, protocols for the reporting of model performance are largely non-existent. As noted elsewhere: “As a community, we have fallen into reliance on measures and procedures for model performance evaluation that say little more than how good or bad the model-to-data comparison is in some ‘average’ sense” (Gupta et al., 2008). Consistent reporting of sets of more informative (than mean squared error or Nash–Sutcliffe efficiency) and properly benchmarked measures of model performance are

necessary to better facilitate the generalization of findings from individual case studies (Mathevet et al., 2006; Perrin et al., 2006; Schaeffli and Gupta, 2007, among others).

It is important to keep in mind that the primary purpose of reporting is to make the information useful to the recipient. The comparative assessment of Blöschl et al. (2013) reported that inconsistency in use and reporting of model performance hampered their investigation. Further, they had to actually approach the authors of the publications to ask for their data and model outputs. To effectively handle such problems, a data repository linked to the papers would have been extremely useful.

4.3 Identification of model structures and parameters

With growing availability of geo-spatial data sets and increases in computational power, there has been significant progress in the development of local-, regional- and continental-scale spatially distributed hydrologic models that simulate hydrological processes at relatively high resolution, and at points that are effectively ungauged (e.g., Carpenter and Georgakakos, 2004; Ivanov et al., 2004; Reed et al., 2004; Koren et al., 2004, Smith et al., 2012). However, while such models make possible detailed support for water resource management (e.g., Blöschl et al., 2008), more work needs to be done on identifiability and transferability of their structures and parameters (e.g., see Beven, 1989, 2002; Grayson et al., 1992; Kirchner, 2006; Samaniego et al., 2010; Andréassian et al., 2012), and on quantifying predictive uncertainty (Beven and Freer, 2001; Doherty and Johnston, 2003; Pokhrel et al., 2008). To establish credibility, such models (and their sub-components) must be properly tested at large numbers of catchments encompassing a wide variety of land-surface and climatic conditions.

5 Perspectives and recommendations

5.1 Overcoming barriers to sharing data

To state the obvious, large-sample hydrology requires large samples of relevant data sets. Most current studies of large-sample catchment hydrology have focused on regional or national scales (e.g., Parajka et al., 2005, 2007; Oudin et al., 2008, 2010; Kumar et al., 2013) and there is a need to extend these to global scale. As computer power and data storage capabilities continue to increase, one might expect that data exchange will also increase. Nonetheless, the PUB synthesis effort (Blöschl et al., 2013), driven largely by the desire to conduct comparative studies, demonstrated clearly that it is indeed possible to compile and jointly analyze large data sets.

Clearly, it will be necessary for the community at large to continue to develop, and vigorously promote, specific policies designed to make hydrology-related data more easily and widely available (Beniston et al., 2012), and this will

require deliberate efforts by a spectrum of organizations, including the International Association of Hydrological Sciences, and the Panta Rhei (Change in Hydrology and Society) initiative. At a more general level it may be necessary to bring about a policy change for data providers and data users. Governments and hydrological services need to be informed about the benefits of shared information for them, and more transparent protocols for the transfer of information may be needed to fully exploit the rich hydrologic data legacy that exists around the world (Viglione et al., 2010). By pointing beyond benchmarking, to the possibility of actually learning new things, we hope that hydrologists and institutions will be convinced of the need to find ways to overcome the economic and legal constraints to building and sharing reference data sets.

Further, as mentioned in Sect. 4.2, easier access to the data and model outputs used in published work will be extremely useful (the Supporting Materials concept now used by several journals is a good step forward). It is common knowledge that the disparity in data characteristics (amount, type, quality, resolution, etc.) from region to region impedes our attempts to make progress in improving process understanding, refining models, improving predictions and reducing uncertainty. In this regard, progress will likely be much faster if large-sample studies include some discussion of the data requirements and availability specific to the various regions, and where and how the models are limited by data. Of course, there will always be regional issues, and issues with interpreting historical data sets.

5.2 Enhancing the link with comparative hydrology

Large-sample studies have traditionally focused on statistical analysis (e.g., to develop regression relationships for flood regionalization or to estimate and transfer model parameters to ungauged basins). With PUB, there has been a move beyond this to focus on the importance of investigating and demonstrating causal processes (Merz and Blöschl, 2008a, b; Carrillo et al., 2011; Troch et al., 2013), by seeking to understand the controls and macro-scale signatures that arise from their actions and interactions (Wagener and Montanari, 2011; Grauso et al., 2008; Sawicz et al., 2011, 2013).

In this regard, the study of comparative hydrology (Falkenmark and Chapman, 1989) seeks to exploit knowledge regarding a much wider array of conditions and processes than can ever be possible with a single model structure, or by studies based on limited numbers of catchments, thereby providing a valuable complement to the detailed investigation of specific catchments. For example, the PUB Synthesis report (Blöschl et al., 2013) revealed significant trends of decreasing performance with increasing aridity, and increasing performance with increasing catchment size and data availability (see Parajka et al., 2013; Salinas et al., 2013), patterns than would have been impossible to detect in any other way. However, comparative studies of

regionalization approaches have resulted in strikingly varied conclusions. One likely reason, though not the only one (see Oudin et al., 2008), may be the use of insufficiently large catchment samples, which would cause the conclusions to be overly dependent on differences in the characteristics represented by each data set.

In fact, large-sample hydrology provides an opportunity for both regionalization and comparative hydrology. While regionalization studies focus on improved hydrological predictions by making use of data from neighboring catchments, in comparative hydrology the focus is on contrasting catchments from different regions to develop generalized understanding about the causes and controls that give rise to these differences. It is noteworthy that the PUB decade of IAHS has seen significant progress in both regionalization approaches and in comparative hydrology as summarized by Blöschl et al. (2013).

Of course, several key issues will need deeper investigation. For example, no clarity has yet emerged regarding which (observable) catchment characteristics are hydrologically relevant to catchment classification, as required for comparative hydrology. Further, how process dominance (and complexity) varies with environment and scale (Skøien et al., 2003), and the role of thresholds, are not well understood. These make it difficult to properly link catchment types to the selection of appropriate process representation (thereby determining model structures). Importantly, classification could help in identifying typical (i.e., normal or representative) catchments to be targeted for intensive investigation, thereby providing a stronger basis for catchment observatory initiatives such as CUAHSI. Meanwhile, the contrast with typical or normal ranges for catchments in a given class will facilitate the detection of outlier catchments similarly deserving of targeted special attention.

5.3 Achieving good performance for the right reasons

Clearly, models structures used in any large-sample investigation should be representative of the diversity of hydrological processes present, so that good performance achieved is for the right reasons (Klemeš, 1986b; Kirchner, 2006; Martinez and Gupta, 2010, 2011). Moreover, the tendency to universally apply a fixed set of assumptions regarding driving mechanisms and process structure can sometimes miss the more important processes in a particular catchment (Savenije, 2009; Halihan et al., 2009; Blöschl et al., 2013).

One obvious approach to model development would be to begin with a highly complex representation and progressively simplify the structure applied to each catchment to achieve a justifiably parsimonious representation. However, as expressed by Bergström (1991), “Going from complex to simpler model structures requires an open mind, because it is frustrating to have to abandon seemingly elegant concepts and theories. It is normally much more stimulating, from an academic point of view, to show significant improvement of

the model performance by increasing complexity.” Following the latter approach, Nash and Sutcliffe (1970) had presented a strategy for model development from the simple to the complex “which may help to avoid this frustration” (Bergström, 1991). This allows one to begin with simple assumptions regarding process structure, identify catchments where these assumptions result in poor performance, and then progressively introduce appropriate complexity justified by diagnostic tests and other evidence. In this regard, the repeated use of a universal model structure as a starting point can aid in the development of the experience necessary to diagnose what does, and does not, work at a specific location.

Ultimately, the goal of working with large catchment sets is to better understand the hydrological cycle. Through a process of designing improved local models, and by understanding the relationships between functional behavior (expressed through model structure) and observable characteristics of the catchment, we should progress towards better comprehension of catchments as systems.

5.4 Estimating model parameters

Structure selection does not, of course, resolve the problem of parameter specification, which is made more complex by the spatial heterogeneities involved in large-sample studies. Fortunately, progress has been made by recognizing that parameter values at specific locations within spatial fields can be linked to observable basin physical characteristics (soil texture, vegetation, topography, etc.) that display strong patterns of organization in space (Abdulla and Lettenmaier, 1997; Fernandez et al., 2000; Hundecha and Bardossy, 2004; Pokhrel et al., 2008; Hundecha et al., 2008; Samaniego et al., 2010; Kumar et al., 2013; among others). For example, some studies have pursued an approach of developing regional relationships that map observable catchment characteristics onto what are sometimes called a priori parameter estimates (Koren et al., 2000; Leavesley et al., 2003; Blöschl, 2005). An important benefit of this approach is to regularize the optimization problem, limiting the degrees of freedom to a small number of regional transfer function coefficients called global-, super-, or hyper-parameters (see Pokhrel et al., 2008; Samaniego et al., 2010).

To obtain robust estimates for these coefficients, the model must be run at a large number of catchments, and across diverse hydro-climatic conditions, so that hypothesis tests about the structures of the regional transfer functions can be conducted. This approach modifies the modeling problem in an interesting way, as one must now evaluate an augmented model hypothesis consisting of both (a) the catchment model structure relating system inputs to state variables and outputs, and (b) the regional transfer function structure and coefficients relating catchment properties to model parameters.

Although the complexity of the hypothesis under investigation is increased, this approach provides statistical advantages by (i) accounting for random variability in catchment

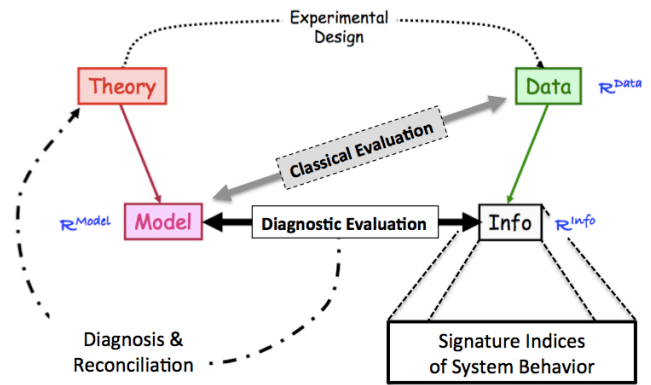


Fig. 4. Classical versus Diagnostic model evaluation. The classical approach compares model simulations directly with the collected data. In the diagnostic approach, *patterns* in the model simulations are compared with corresponding *patterns* in the data (figure based on ideas from Gupta et al., 2008).

properties that tends to diminish their correlation with lumped catchment-scale properties (Kling and Gupta, 2009); (ii) reducing the biasing effects of data noise due to the damping effects of larger sample sizes; and (iii) improving identifiability due to increased diversity of hydro-climatic conditions covered. The recent multi-scale parameter regionalization approach (Samaniego et al., 2010; Kumar et al., 2013) suggests that the approach is robust, and can facilitate the transfer of model hypotheses across spatial domains. Further strength could be achieved by constraining the model using information about hydrological dynamics provided by soil moisture, snow cover, etc. (Parajka and Blöschl, 2008; Parajka et al., 2009).

5.5 Testing hypotheses and assessing reliability

Ultimately, for models to be demonstrably robust, they must pass the kind of crash testing proposed by Linsley (1982), Klemeš (1986a), Andréassian et al. (2009), and Coron et al. (2012), among others. While large-sample studies offer a unique opportunity to test hypotheses that are based in process understanding, the aggregate performance indices commonly used in hydrology for performance assessment are poorly benchmarked (Legates and McCabe, 1999; Seibert, 2001; Schaeffli and Gupta, 2007; Parajka et al., 2013) and not easily related to catchment properties and functioning (Gupta et al., 2008). They are, therefore, of limited usefulness in large-sample assessments. A key to better use of large-sample studies is to find ways to relate model performance and predictive uncertainty (and therefore model adequacy) to catchment structure and function, thereby providing insight into which processes the model is incapable of describing well (Fig. 1) and moving beyond the commonplace approach of tuning the model to compensate for model structural adequacy problems (Gupta et al., 2012).

In this regard, it will be helpful to shift the focus of model evaluation away from data fitting towards an emphasis on reproduction of diagnostic signatures (Fig. 4) – a move away from hydrograph mimicry towards model fidelity (Vogel and Sankarasubramanian, 2003; Yadav et al., 2007; Gupta et al., 2008; Blöschl et al., 2008; Martinez and Gupta, 2011; Clark et al., 2011; Koster and Mahanama, 2012; Castiglioni et al., 2010; etc.) – the idea being to make it harder to win the game through calibration (in the sense of model tuning) and to instead seek answers that are correct for the right reasons (Kirchner, 2006). Therefore, further progress in diagnostic analysis (Yilmaz et al., 2008; Martinez and Gupta, 2010, 2011; McMillan et al., 2011), improved understanding of physical land-surface constraints (Koster and Mahanama, 2012), and improved assessments of model structural adequacy (Gupta et al., 2012) and uncertainty (Montanari et al., 2009; Montanari, 2011; Montanari and Koutsoyiannis, 2012) are required. All of these areas depend on increased analytical sophistication, and will undoubtedly benefit from the breadth of information contained in large-sample data sets.

5.6 Coping with variability and change

The non-stationarities underlying changing climate and catchment conditions make the problem of estimating hydrological fluxes and predicting catchment response more difficult. In such cases, it becomes important to distinguish between situations that are predictable from those that are not (Blöschl and Montanari, 2010; Kumar, 2011). It is now common to use scenario analysis as a way of assessing the impacts of future change on hydrological response (Mahmoud et al., 2009, 2011; among many others). However, investigation using large-sample data sets can facilitate a much wider range of analyses.

For example, large-sample data sets were used by Merz et al. (2011) and Coron et al. (2012) to investigate changes in estimated model parameters associated with changes in climate, and by Ter Braak and Prentice (1988) to use spatial gradients as a surrogate for temporal gradients in investigations of change. Similarly, Peel and Blöschl (2011) suggested that changing climate could cause hydrological processes in one catchment to become similar to those in other catchments currently experiencing conditions similar to the target climate, thereby providing a basis for understanding the effects of change. Ultimately, these analyses must exploit the information provided by data other than runoff, including hydrogeological information, soil moisture from remote sensing products (Parajka et al., 2006), snow characteristics (e.g., Blöschl and Kirnbauer, 1991, 1992), etc.

5.7 Trading depth of analysis for sample size

Finally, while the benefits of large-sample hydrology are many and important, they come at a cost. When large numbers of catchments are analyzed, all procedures for selecting

model structures and estimating parameters must be automated. This makes it difficult to attend to clues that might otherwise be provided by the assessment of local knowledge (whether hard or soft data), or by process understanding that can be gained from field trips. In this regard, soft information on catchment functioning can be as valuable as hard data for improving understanding about catchment functioning (Seibert and McDonnell, 2002). Ultimately, there is no free lunch, and investigations that exploit the benefits of large catchment sample sizes (providing breadth) must be complemented by detailed investigations at specifically targeted catchments (providing depth).

6 In conclusion

This paper has discussed the need to actively promote and pursue the use of a large catchment sample approach to modeling the rainfall–runoff process, thereby balancing depth with breadth. In either case, the need to understand hydrological change will require that large-sample investigations be guided by process understanding, with statistical analysis playing the important supporting roles of (a) capturing the summary effects of various controls, including feedbacks across processes and scales, and (b) detecting spatial and temporal patterns that will need to be properly explained. Only then can we expect to achieve improved predictability and the ability to extrapolate to new situations (Kumar, 2011).

We hope that this paper serves the function of provoking further discussion, and will promote what could be an important theme for Panta Rhei, the upcoming IAHS Scientific Decade. In this regard, as suggested by reviewer Sopan Patil, we end with a call for increasing the diversity of efforts in large-sample hydrology (beyond the focus on catchment modeling taken by this paper).

Supplementary material related to this article is available online at <http://www.hydrol-earth-syst-sci.net/18/463/2014/hess-18-463-2014-supplement.pdf>.

Acknowledgements. This paper was partially developed in the context of the Panta Rhei research initiative of the International Association of Hydrological Sciences. We wish to express our appreciation to the Editor, Murugesu Sivapalan, and the several reviewers, Bethanna Jackson, Hillary McMillan, Dingbao Wang, Evan Coopersmith, Sopan Patil, Manfred Ostrowski and Thorsten Wagener, plus one additional anonymous reviewer, for their detailed and constructive comments. Their many remarks and suggestions helped to streamline and improve the presentation of our opinion paper. The first author acknowledges partial support by a grant from the Co-operative Research Programme (Trade and Agriculture) of the Organization for European Co-Operation and Development (OECD), and from the Australian Research Council

through the Centre of Excellence for Climate System Science (grant number CE110001028). The third author acknowledges funding from the Austrian Science Fund (projects W1219-N22 and P 23723-N21). Olivier Delaigue is thanked for providing Fig. 2.

Edited by: M. Sivapalan

References

- Abdulla, F. A., and Lettenmaier, D. P.: Development of regional parameter estimation equations for a macroscale hydrologic model, *J. Hydrol.*, 197, 230–257, 1997.
- Andréassian, V., Hall, A., Chahinian, N., and Schaake, J.: Introduction and Synthesis: Why should hydrologists work on a large number of basin data sets?, in: IAHS Publication no. 307, Large sample basin experiments for hydrological model parameterization: Results of the Model Parameter Experiment – MOPEX, Paris, 1–5, 2006.
- Andréassian, V., Lerat, J., Loumagne, C., Mathevet, T., Michel, C., Oudin, L., and Perrin, C.: What is really undermining hydrologic science today?, *Hydrol. Process.*, 21, 2819–2822, doi:10.1002/hyp.6854, 2007.
- Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions “Crash tests for a standardized evaluation of hydrological models”, *Hydrol. Earth Syst. Sci.*, 13, 1757–1764, doi:10.5194/hess-13-1757-2009, 2009.
- Andréassian, V., Perrin, C., Parent, E., and Bardossy, A.: Editorial – The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?, *Hydrolog. Sci. J.*, 55, 849–856, doi:10.1080/02626667.2010.506050, 2010.
- Andréassian, V., Le Moine, N., Perrin, C., Ramos, M. H., Oudin, L., Mathevet, T., Lerat, J., and Berthet, L.: All that glitters is not gold: the case of calibrating hydrological models, *Hydrol. Process.*, 26, 2206–2210, doi:10.1002/hyp.9264, 2012.
- Beniston, M., Stoffel, M., Harding, R., Kernan, M., Ludwig, R., Moors, E., Samuels, P., and Tockner, K.: Obstacles to data access for research related to climate and water: Implications for science and EU policy-making, *Enviro. Sci. Policy*, 17, 41–48, 2012.
- Bergström, S.: Principles and confidence in hydrological modelling, *Nord. Hydrol.*, 22, 123–136, 1991.
- Beven, K.: Changing ideas in hydrology – The case of physically-based models, *J. Hydrol.*, 105, 157–172, 1989.
- Beven, K.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, *Hydrol. Process.*, 16, 189–206, 2002.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems, *J. Hydrol.*, 249, 11–29, 2001.
- Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, *Hydrol. Earth Syst. Sci.*, 4, 203–213, doi:10.5194/hess-4-203-2000, 2000.
- Blöschl, G.: Rainfall-runoff modeling of ungauged catchments, in: *Encyclopedia of Hydrological Sciences*, edited by: Anderson, M. G., John Wiley & Sons, Chichester, 2061–2080, 2005.
- Blöschl, G. and Kirnbauer, R.: Point snowmelt models with different degrees of complexity – internal processes, *J. Hydrol.*, 129, 127–147, 1991.
- Blöschl, G. and Kirnbauer, R.: An analysis of snow cover patterns in a small Alpine catchment, *Hydrol. Process.*, 6, 99–109, 1992.
- Blöschl, G. and Montanari, A.: Climate change impacts – throwing the dice?, *Hydrol. Process.*, 24, 374–381, 2010.
- Blöschl, G., Reszler, C., and Komma, J.: A spatially distributed flash flood forecasting model, *Environ. Modell. Softw.*, 23, 464–478, doi:10.1016/j.envsoft.2007.06.010, 2008.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: *Runoff Prediction in Ungauged Basins, Synthesis across Processes, Places and Scales*, Cambridge University Press, Cambridge, UK, 465 pp., 2013.
- Boldetti, G., Riffard, M., Andréassian, V., and Oudin, L.: Dataset cleansing practices and hydrological regionalization: is there any valuable information among outliers?, *Hydrolog. Sci. J.*, 55, 941–951, 2010.
- Boughton, W.: A review of Australian model parameterization studies using large basin samples, IAHS Publication no. 307, IAHS, Wallingford, UK, 126–131, 2006.
- Bourgin, F., Andréassian, V., Perrin, C., and Oudin, L.: Transferring model uncertainty estimates from gaged to ungauged catchments, submitted, 2013.
- Carpenter, T. M. and Georgakakos, K. P.: Continuous streamflow simulation with the HRC-DHM distributed hydrologic model, *J. Hydrol.*, 298, 61–79, 2004.
- Carrillo, G., Troch, P. A., Sivapalan, M., Wagener, T., Harman, C., and Sawicz, K.: Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient, *Hydrol. Earth Syst. Sci.*, 15, 3411–3430, doi:10.5194/hess-15-3411-2011, 2011.
- Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A.: Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, *Adv. Water Resour.*, 33, 1235–1242, doi:10.1016/j.advwatres.2010.04.009, 2010.
- Chahinian, N., Mathevet, T., Habets, F., and Andréassian, V.: The MOPEX 2004 French database: main hydrological and morphological characteristics, IAHS, Wallingford, UK, 29–40, 2006.
- Clark, M. P., Kavetski, D., and Fencica, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, doi:10.1029/2010WR009827, 2011.
- Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments, *Water Resour. Res.*, 48, W05552, doi:10.1029/2011WR011721, 2012.
- Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWPb2: Multimodel Analysis and Implications for Our Perception of the Land Surface, *B. Am. Meteorol. Soc.*, 87, 1381–1397, 2006.
- Doherty, J. and Johnston, J. M.: Methodologies for calibration and predictive analysis of a watershed model, *J. Am. Water Resour. Assoc.*, 39, 251–265, 2003.
- Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results

- from the second and third workshops, *J. Hydrol.*, 320, 3–17, doi:10.1016/j.jhydrol.2005.07.031, 2006.
- Egbuniwe, N. and Todd, D. K.: Application of the Stanford Watershed Model to Nigerian watersheds, *Water Resour. Bull.*, 12, 449–460, 1976.
- Ehret, U., Blöschl, G., Bogaard, T. A., Gelfan, A. N., Gupta, H. V., Harman, C., Kleidon, A., Scherer, U., Schymanski, S., Sivapalan, M., Wang, D., Weijs, S. V., Wagener, T., Bierkens, M. F. P., Di Baldassarre, G., Parajka, J., van Beek, L. P. H., van Griensven, A., Westhoff, M., Winsemius, H. C., and Zehe, E.: Advancing Hydrology to deal with Predictions Under Change, *Hydrolog. Sci. J.*, 58, 1198–1255, 2013
- Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, *Hydrol. Earth Syst. Sci.*, 17, 1893–1912, doi:10.5194/hess-17-1893-2013, 2013.
- Falkenmark, M. and Chapman, T. (Eds.): *Comparative Hydrology*, UNESCO, Paris, 1989.
- Fernandez, W., Vogel, R. M., and Sankarasubramanian, A.: Regional calibration of a watershed model, *Hydrolog. Sci. J.*, 45, 689–707, 2000.
- Freebairn, J. and Zillman, J.: Funding meteorological services, *Meteorol. Appl.*, 9, 45–54, 2002.
- Goodall, J. L., Horsburgh, J. S., Whiteaker, T. L., Maidment, D. R., and Zaslavsky, I.: A first approach to web services for the National Water Information System, *Environ. Modell. Softw.*, 23, 404–411, doi:10.1016/j.envsoft.2007.01.005, 2008.
- Göttinger, J. and Bardossy, A.: Comparison of four regionalisation methods for a distributed hydrological model, *J. Hydrol.*, 333, 374–384, doi:10.1016/j.jhydrol.2006.09.008, 2007.
- Grauso, S., Fattoruso, G., Crocetti, C., and Montanari, A.: Estimating the suspended sediment yield in a river network by means of geomorphic parameters and regression relationships, *Hydrol. Earth Syst. Sci.*, 12, 177–191, doi:10.5194/hess-12-177-2008, 2008.
- Grayson, R. B., Moore, I. D., and McMahon, T. A.: Physically-based modelling, 2. Is the concept realistic, *Water Resour. Res.*, 26, 2659–2666, 1992.
- Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Towards important Calibration of Hydrologic Models: Multiple and Non-Commensurable Measures of Information, *Water Resour. Res.*, 34, 751–763, 1998.
- Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Process.*, 22, 3802–3813, 2008.
- Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a Comprehensive Assessment of Model Structural Adequacy, *Opinion Paper*, *Water Resour. Res.*, 48, W08301, doi:10.1029/2011WR011044, 2012.
- Halihan, T., Mouri, S., and Puckette, J.: Evaluation of Fracture Properties of the Arbuckle Simpson Aquifer, report Oklahoma State University, www.owrb.ok.gov/studies/groundwater/arbuckle_simpson/pdf/2009_Reports/EvaluationFracturePropertiesArbuckleSimpson_Halihan.pdf, Oklahoma State University, 2009.
- Hardison, C. H.: Generalized skew coefficients of annual floods in the United States and their application, *Water Resour. Res.*, 10, 745–752, doi:10.1029/WR010i004p00745, 1974.
- Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: An appraisal of the regional flood frequency procedure in the UK Flood Studies Report, *Hydrolog. Sci. J.*, 30, 85–109, doi:10.1080/02626668509490973, 1985.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) – a review, *Hydrolog. Sci. J.*, 58, 1198–1255, doi:10.1080/02626667.2013.803183, 2013.
- Hughes, D. A.: A review of 40 years of hydrological science and practice in southern Africa using the Pitman rainfall–runoff model, *J. Hydrol.*, 501, 111–124, doi:10.1016/j.jhydrol.2013.07.043, 2013.
- Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J. Hydrol.*, 292, 281–295, doi:10.1016/j.jhydrol.2004.01.002, 2004.
- Hundecha, Y., Ouarda, T., and Bardossy, A.: Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the “spatial” structures of the parameters within a canonical physiographic-climatic space, *Water Resour. Res.*, 44, W01427, doi:10.1029/2006WR005439, 2008.
- Ivanov, V. Y., Vivoni, E. R., Bras, R. L., and Entekhabi, D.: Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach, *J. Hydrol.*, 298, 80–111, doi:10.1016/j.jhydrol.2004.03.041, 2004.
- James, L. D.: Hydrologic modeling, parameter estimation, and watershed characteristics, *J. Hydrol.*, 17, 283–307, 1972.
- Jones, J. R., Beall, R. M., and Giusti, E. V.: International cooperation in water resources, *GeoJournal*, 3, 481–487, 1979.
- Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362, 2006.
- Klemeš, V.: Operational testing of hydrological simulation models, *Hydrolog. Sci. J.*, 31, 13–24, 1986a.
- Klemeš, V.: Dilettantism in hydrology: transition or destiny?, *Water Resour. Res.*, 22, 177S–188S, 1986b.
- Kling, H. and Gupta, H.: On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, *J. Hydrol.*, 373, 337–351, 2009.
- Koren, V., Reed, S., Smith, M., Zhang, Z., and Seo, D. J.: Hydrology Laboratory Research Modeling System (HL-RMS) of the US National Weather Service, *J. Hydrol.*, 291, 297–318, doi:10.1016/j.jhydrol.2003.12.039, 2004.
- Koren, V. I., Smith, M., Wang, D., and Zhang, Z.: Use of soil property data in the derivation of conceptual rainfall runoff model parameters, *Proceedings of the 15th Conference on Hydrology*, American Meteorological Society, Long Beach, California, 103–106, 2000.
- Koster, R. D. and Mahanama, S. P. P.: Land Surface Controls on Hydroclimatic Means and Variability, *J. Hydrometeorol.*, 13, 1604–1620, 2012.

- Kovács, G.: Proposal to construct a coordinating matrix for comparative hydrology, *Hydrolog. Sci. J.*, 29, 435–443, doi:10.1080/02626668409490961, 1984.
- Kruger, A., Krajewski, W. F., Domaszczynski, P., and Smith, J. A.: Hydro-NEXRAD: Metadata computation and use, *J. Hydroinform.*, 13, 267–276, 2011.
- Kumar, P.: Typology of hydrologic predictability, *Water Resour. Res.*, 47, W00H05, doi:10.1029/2010WR009769, 2011.
- Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, *Water Resour. Res.*, 49, 360–379, doi:10.1029/2012wr012195, 2013.
- Leavesley, G. H., Hay, L. E., Viger, R. J., and Markstrom, S. L.: Use of a priori parameter estimation methods to constrain calibration of distributed parameter models, in: *Calibration of Watershed Models*, Water, Science and Application 6, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., American Geophysical Union, 255–266, 2003.
- Legates, D. R. and McCabe, G. J.: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, 35, 233–241, 1999.
- Le Moine, N., Andréassian, V., Perrin, C., and Michel, C.: “Outlier” catchments: what can we learn from them in terms of prediction uncertainty in rainfall-runoff modelling?, *IAHS Publ.*, 313, 195–203, 2007.
- Lichty, R. W. and Karlinger, M. R.: Climate factor for small-basin flood frequency, *J. Am. Water Resour. Assoc.*, 26, 577–586, doi:10.1111/j.1752-1688.1990.tb01395.x, 1990.
- Linsley, R. K.: Rainfall–runoff models – An overview, in: *Rainfall–runoff relationship*, edited by: Singh, V. P., Water Resources Publications, Highlands Ranch, Colorado, USA, 3–22, 1982.
- Loague, K. M. and Freeze, R. A.: A comparison of rainfall–runoff modeling techniques on small upland catchments, *Water Resour. Res.*, 21, 229–248, 1985.
- Magette, W. L., Shanholtz, V. O., and Carr, J. C.: Estimating selected parameters for the Kentucky watershed model from watershed characteristics, *Water Resour. Res.*, 12, 472–476, 1976.
- Mahmoud, M., Liu, Y., 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., and Winter, L.: A formal framework for scenario development in support of environmental decision-making, *Environ. Modell. Softw.*, 24, 798–808, doi:10.1016/j.envsoft.2008.11.010, 2009.
- Mahmoud, M., Gupta, H. V., and Rajagopal, S.: Scenario development for water resources planning and watershed management: Methodology and semi-arid region case study, *Environ. Modell. Softw.*, 26, 873–885, doi:10.1016/j.envsoft.2011.02.003, 2011.
- Martinez, G. F. and Gupta, H. V.: Toward improved identification of hydrological models: a diagnostic evaluation of the “abcd” monthly water balance model for the conterminous United States, *Water Resour. Res.*, 46, W08507, doi:10.1029/2009WR008294, 2010.
- Martinez, G. F. and Gupta, H. V.: Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States, *Water Resour. Res.*, 47, W12540, doi:10.1029/2011WR011229, 2011.
- Mathevet, T., Michel, C., Andréassian, V., and Perrin, C.: A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins, in: *Large sample basin experiments for hydrological model parameterisation: Results of the Model Parameter Experiment – MOPEX*, edited by: Andréassian, V., Hall, A., Chahinian, N., and Schaake, J., IAHS Red Books Series no. 307, IAHS, Wallingford, UK, 211–219, 2006.
- McDonnell, J. J. and Woods, R.: On the need for catchment classification, *J. Hydrol.*, 299, 2–3, 2004.
- McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., and Woods, R. A.: Hydrological field data from a modeller’s perspective: Part 1. Diagnostic tests for model structure, *Hydrol. Process.*, 25, 511–522, doi:10.1002/hyp.7841, 2011.
- Mein, R. G. and Brown, B. M.: Sensitivity of optimized parameters in watershed models, *Water Resour. Res.*, 14, 299–303, 1978.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007WR006744, 2008a.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 2. Combining data evidence, *Water Resour. Res.*, 44, W08433, doi:10.1029/2007WR006745, 2008b.
- Merz, R., Blöschl, G., and Parajka, J.: Regionalisation methods in rainfall-runoff modelling using large samples, IAHS publication no. 307, IAHS, Wallingford, UK, 117–125, 2006.
- Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, 47, W02531, doi:10.1029/2010WR009505, 2011.
- Michel, C., Perrin, C., Andréassian, V., Oudin, L., and Mathevet, T.: Has basin scale modelling advanced far beyond empiricism?, IAHS Publication no. 307, IAHS, Wallingford, UK, 108–116, 2006.
- Montanari, A.: Uncertainty of Hydrological Predictions, in: *Treatise on Water Science*, vol. 2, edited by: Wilderer, P., Academic Press, Oxford, 459–478, 2011.
- Montanari, A. and Koutsoyiannis, D.: A blueprint for process-based modeling of uncertain hydrological systems, *Water Resour. Res.*, 48, W09555, doi:10.1029/2011WR011412, 2012.
- Montanari, A., Shoemaker, C. A., and van de Giesen, N.: Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and challenges, *Water Resour. Res.*, 45, W00B00, doi:10.1029/2009wr008471, 2009.
- Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaeffli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: “Panta Rhei-Everything Flows”: Change in hydrology and society – The IAHS Scientific Decade 2013–2022, *Hydrolog. Sci. J.*, 58, 1256–1275, doi:10.1080/02626667.2013.809088, 2013.
- Naef, F.: Can we model the rainfall-runoff process today?, *Hydrol. Sci. Bull.*, 26, 281–289, 1981.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models. Part I - A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.

- NOAA: Office of Hydrologic Development Hydrology Laboratory Strategic Science Plan, http://www.nws.noaa.gov/oh/src/docs/Strategic_SciencePlan2007-Final.pdf, 2007.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity and ungauged catchments: confrontation on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240, 2008.
- Oudin, L., Kay, A. L., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, 46, W11558, doi:10.1029/2009WR008887, 2010.
- Parajka, J. and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, *J. Hydrol.*, 358, 240–258, doi:10.1016/j.jhydrol.2008.06.006, 2008.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.
- Parajka, J., Naemi, V., Blöschl, G., Wagner, W., Merz, R., and Scipal, K.: Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, *Hydrol. Earth Syst. Sci.*, 10, 353–368, doi:10.5194/hess-10-353-2006, 2006.
- Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, *Hydrol. Process.*, 21, 435–446, doi:10.1002/hyp.6253, 2007.
- Parajka, J., Naemi, V., Blöschl, G., and Komma, J.: Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria, *Hydrol. Earth Syst. Sci.*, 13, 259–271, doi:10.5194/hess-13-259-2009, 2009.
- Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, 17, 1783–1795, doi:10.5194/hess-17-1783-2013, 2013.
- Peel, M. C. and Blöschl, G.: Hydrological modelling in a changing world, *Prog. Phys. Geogr.*, 35, 249–261, 2011.
- Peel, M. C., Chiew, F. H. S., Western, A. W., and McMahon, T. A.: Extension of Unimpaired Monthly Streamflow Data and Regionalisation of Parameter Values to Estimate Streamflow in Ungauged Catchments, National Land and Water Resources Audit, Theme 1 – Water Availability, Centre for Environmental Applied Hydrology, The University of Melbourne, Melbourne, p. 37, 2000.
- Perrin, C., Andréassian, V., and Michel, C.: Simple benchmark models as a basis for criteria of model efficiency, *Archiv für Hydrobiologie Supplement* 161/1–2, Large Rivers, 17, 221–244, 2006.
- Pirt, J. and Bramley, E. A.: The application of simple moisture accounting models to ungauged catchments, *J. Inst. Water Eng. Sci.*, 39, 169–177, 1985.
- Pokhrel, P., Gupta, H. V., and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, *Water Resour. Res.*, 44, W12419, doi:10.1029/2007WR006615, 2008.
- Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., and Seo, D. J.: Overall distributed model intercomparison project results, *J. Hydrol.*, 298, 27–60, 2004.
- Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, *Hydrol. Earth Syst. Sci.*, 17, 2637–2652, doi:10.5194/hess-17-2637-2013, 2013.
- Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a gridbased hydrologic model at the mesoscale, *Water Resour. Res.*, 46, W05523, doi:10.1029/2008WR007327, 2010.
- Savenije, H. H. G.: HESS Opinions “The art of hydrology”*, *Hydrol. Earth Syst. Sci.*, 13, 157–161, doi:10.5194/hess-13-157-2009, 2009.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.
- Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., and Carrillo, G.: Technical Note: Characterizing hydrologic change through catchment classification, *Hydrol. Earth Syst. Sci. Discuss.*, 10, 6599–6627, doi:10.5194/hessd-10-6599-2013, 2013.
- Schaake, J. and Duan, Q. Y.: The model parameter estimation experiment (MOPEX) – Preface, *J. Hydrol.*, 320, 1–2, 2006.
- Schaake, J., Duan, Q., Koren, V., and Hall, A.: Toward improvement parameter-estimation of land surface hydrology models through the Model Parameter Estimation Experiment (MOPEX), *IAHS-AISH Publ.*, 270, 91–98, 2001.
- Schaeffli, B. and Gupta, H. V.: Do Nash values have value?, *Hydrol. Process.*, 21, 2075–2080, 2007.
- Seibert, J.: On the need for benchmarks in hydrological modelling, *Hydrol. Processes*, 15, 1063–1064, 2001.
- Seibert, J., and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, 38, 1241, doi:10.1029/2001WR000978, 2002.
- Sittner, W. T.: WMO project intercomparison of conceptual models used in hydrological forecasting, *Hydrolog. Sci. J.*, 21, 203–213, 1976.
- Sivapalan, M.: Prediction in ungauged basins: a grand challenge for theoretical hydrology, *Hydrol. Process.*, 17, 3163–3170, 2003.
- Skøien, J. O. and Blöschl, G.: Spatiotemporal topological kriging of runoff time series, *Water Resour. Res.*, 43, W09419, doi:10.1029/2006WR005760, 2007.
- Skøien, J. O., Blöschl, G., and Western, A. W.: Characteristic space scales and timescales in hydrology, *Water Resour. Res.*, 39, 1304, doi:10.1029/2002WR001736, 2003.
- Smith, M. B., Seo, D. J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, *J. Hydrol.*, 298, 4–26, 2004.
- Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E. A., and Cosgrove, B. A.: The distributed model intercomparison project – Phase 2: Motivation and design of the Oklahoma experiments, *J. Hydrol.*, 418–419, 3–16, 2012.

- Srikanthan, R. and Goodspeed, M. J.: Regionalization of conceptual model parameters for meso-scale catchments in the Hunter Valley, Hydrology and Water Resources Symposium 1988, ANU, Canberra, 85–90, 1988.
- Ter Braak, C. J. F. and Prentice, I. C.: A theory of gradient analysis, *Adv. Ecol. Res.*, 18, 271–317, 1988.
- Thompson, S. E., Harman, C. J., Konings, A. G., Sivapalan, M., Neal, A., and Troch, P. A.: Comparative hydrology across AmeriFlux sites: The variable roles of climate, vegetation, and groundwater, *Water Resour. Res.*, 47, W00J07, doi:10.1029/2010WR009797, 2011.
- Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T., and Sawicz, K.: Climate-vegetation-soil interactions and long-term hydrologic partitioning: signatures of catchment co-evolution, *Hydrol. Earth Syst. Sci.*, 17, 2209–2217, doi:10.5194/hess-17-2209-2013, 2013.
- Uhlemann, S., Bertelmann, R., and Merz, B.: Data expansion: the potential of grey literature for understanding floods, *Hydrol. Earth Syst. Sci.*, 17, 895–911, doi:10.5194/hess-17-895-2013, 2013.
- van der Ent, R. J. and Savenije, H. H. G.: Length and time scales of atmospheric moisture recycling, *Atmos. Chem. Phys.*, 11, 1853–1863, doi:10.5194/acp-11-1853-2011, 2011.
- Viglione, A., Borga, M., Balabanis, P., and Blöschl, G.: Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy, *J. Hydrol.*, 394, 63–77, 2010.
- Vogel, R. M. and Sankarasubramanian, A.: Validation of a watershed model without calibration, *Water Resour. Res.*, 39, 1292, doi:10.1029/2002WR001940, 2003.
- Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, *Water Resour. Res.*, 47, W06301, doi:10.1029/2010WR009469, 2011.
- Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, *Geogr. Compass*, 1, 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.
- Weeks, W. D. and Ashkanasy, N. M.: Regional parameters for the Sacramento model: a case study, *Trans. Inst. Eng. Aust.*, CE27, 305–313, 1985.
- Weeks, W. D. and Hebbert, R. H. B.: A comparison of rainfall-runoff models, *Nord. Hydrol.*, 11, 7–24, 1980.
- Williams, C. A., Reichstein, M., Buchmann, N., Baldocchi, D., Beer, C., Schwalm, C., Wohlfahrt, G., Hasler, N., Bernhofer, C., Foken, T., Papale, D., Schymanski, S., and Schaefer, K.: Climate and vegetation controls on the surface water balance: Synthesis of evapotranspiration measured across a global network of flux towers, *Water Resour. Res.*, 48, W06523, doi:10.1029/2011WR011586, 2012.
- Winter, T. C.: The concept of hydrologic landscapes, *J. Am. Water Resour. Assoc.*, 37, 335–349, 2001.
- WMO: Intercomparison of conceptual models used in operational hydrological forecasting, Operational Hydrology Report no. 7, WMO no. 429, World Meteorological Organization, Geneva, Switzerland, 1975.
- WMO: Intercomparison of models of snowmelt runoff, Operational Hydrology Report no. 23, WMO no. 646, World Meteorological Organization, Geneva, Switzerland, 1986.
- WMO: Simulated real-time intercomparison of hydrological models, Operational Hydrology Report no. 38, WMO 779, World Meteorological Organization, Geneva, Switzerland, 1992.
- Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30, 1756–1774, 2007.
- Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resour. Res.*, 44, W09417, doi:10.1029/2007WR006716, 2008.