HESS Opinions[AB1]

"Advocating Process Modeling and De-Emphasizing Parameter Estimation"

A. Bahremand

Watershed Management Department, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

Correspondence to: A. Bahremand (Abdolreza.Bahremand@yahoo.com)

Abstract

Since its origins as an engineering discipline, with its widespread use of '*black box*' (empirical) modelling approaches, hydrology has evolved into a scientific discipline that seeks a more '*white box*' (physics-based) modelling approach to solving problems such as the description and simulation of the rainfall-runoff responses of a watershed. There has been much recent debate regarding the future of the hydrological sciences, and several publications have voiced opinions on this subject. This opinion paper seeks to comment and expand upon some recent publications that have advocated an increased focus on process-based modelling while de-emphasizing the focus on detailed attention to parameter estimation. In particular, it offers a perspective that emphasizes a more hydraulic (more physics-based and less empirical) [AB2] approach to development and implementation of hydrological models.

1 Introduction

There has been a recent call in several notable publications for a new focus to be brought to the hydrological sciences. As an example, *Montanari et al. (2015)* stressed the need for new vision, to help drive new theories, new methods and "new thinking". This comes at a time when enhanced computational power and sophisticated monitoring techniques now enable hydrologists to pursue deeper investigations of hydrologic processes, and to thereby simulate watershed hydrology in ever more detail.

It is my opinion that we need to take a broader look at the practices we bring to hydrological modelling. My experience suggests that we too often allow ourselves to become mired in relatively minor problems, and thereby fail to notice some of the more major ones. For example, do we not tend to become over-focused on estimating parameter values by “optimization”, and should we not instead devote more of our focus to improve the models that represent the underlying system processes? Is it not possible to conduct model evaluation (as a support for model building) in a much more intellectually satisfying manner? This paper, while commenting on and referring to some related publications, seeks to promote discussion of such questions and advocates the need for enhanced focus on understanding and representing hydrological processes accurately, so as to improve our conceptual understanding and even our hydrological perceptions.

2 On model parameterization and the need for parameter optimization

In a recent debate on the future of hydrological sciences, and in the context of a discussion of modeled process parameterization and parameter estimation, *Gupta and Nearing (2014)* state that "*we suggest that much can be gained by focusing more directly on the a priori role of Process Modeling (particularly System Architecture) while de-emphasizing detailed System Parameterizations*". Soon after, *Gharari et al. (2014)* presented a practical and methodical demonstration that the need for model calibration (optimization of parameter values) can be dramatically reduced (and even avoided) by the judicious imposition of (both general and site-specific) relational parameter and process constraints onto our models. They report that doing so can significantly improve the results while reducing simulation uncertainty.

The arguments and demonstration mentioned above are recent contributions to a long-standing perspective held by others in the hydrological community. *Bergstrom (2006)*, for example, based on his experience with the HBV model as a solution for prediction in ungauged basins, mentions three possible ways that runoff in rivers can be estimated in the absence of directly available data. "*The first was to simply use information from neighboring rivers through statistical methods. The second option was to get so much experience with a conceptual model that we can map the optimum values of its parameters, or relate them to catchment characteristics. The third* *was to use a model that is so physically correct that it does not need calibration at all*" (*Bergstrom, 2006*).

My own experience, based on working with a physics- and GIS-based fully distributed hydrologic model called WetSpa, is similar to the second aforementioned option proposed by *Bergstrom (2006)*, and resonates with the "*limited need for calibration*" shown so nicely by *Gharari et al. (2014)* (see also *Hrachowitz et al. 2014*). I have found that the need for parameter calibration can be dramatically reduced simply by avoiding the now-common "*trial and error*" strategy of search by optimization, and proceeding instead by a) beginning with some reasonable initial values derived based on known catchment characteristics, b)[AB3] some trial and error to refine the reasonable initial values, and c) proceeding to imposing some meaningful and sensible constraints and parameter relational rules. I find that, much of the time, excellent parameter values (and hence model performance) can be obtained in only a few attempts and without considerable effort. With some degree of practice, and after gaining some understanding about how the hydrological processes are represented in the model and how the parameters relate to observable or conceptual catchment characteristics, the process of model calibration is eased to such an extent that it would imply that the model needs no parameter calibration but only a kind of parameter "*allocation*" (i.e., a logic-based specification); I will discuss parameter allocation in detail later in this paper.

According to *Beven (2000, 2006, 2011)* and *McDonnell and Beven (2014)* the importance of uniqueness of place and the limitations of hydrological data can, in most cases, make parameter allocation rather difficult, and so we should consider the limitations of current concepts. As[AB4] mentioned by *Beven* in his referee comment, in practice we are both model and data limited, and even a perfect model will be limited by inconsistencies in the calibration and prediction data (e.g. *Beven and Smith, 2014*) – so that the success or failure of a model run with a priori parameter estimates might depend more on the (unknown) errors in the data than on whether the model is a realistic representation of the processes.

However, the work of Bergstrom with the HBV model, and more recently *Semenova and Beven* [AB5]*(2015)* seems to suggest otherwise (although note that Beven has a different opinion in this regards, as discussed briefly in their paper; see also Beven's equifinality thesis in *Beven, 2006b*). The work of the St. Petersburg [AB6]modeling team on a deterministic distributed process-based model of runoff formation processes named "hydrograph model" is closely in line with what is described for parameter estimation in this opinion paper (*Vinogradov, 1990, Vinogradov et al. 2011, Semenova et al. 2013 and 2015, Lebedeva et al. 2014*). In their approach, they "do not accept calibration in the form of automated procedure of parameter estimation", and "assume its common application to be one of the main barriers in development of modern hydrological modeling" (www.hydrograph-model.ru).

It seems, in fact, that it may often be possible to arrive at parameter values through a process of reasoning and white box modeling, rather than by the inefficient and poorly informed search procedures involved in trial-and-error or black box efforts. As another example of the use of knowledge from processes to constrain parameters in a physically based, spatially distributed model, I note the TOPKAPI modeling work of *Ragettli and Pellicciotti (2012*) in a glacier-dominated basin; their report includes an evaluation of the transferability of such parameters in time and space.

To estimate the parameters of a spatially distributed flash flood model, *Blosch et al. (2008)* have emphasized understanding the model behavior over formal calibration. Similarly, *Merz and Blosch (2008a, 2008b)* and *Viglione et al (2013)* provide good examples of the use of hydrological reasoning to obtain more informed estimates of flood frequencies, and *Hingray et al. (2010)* present a signature-based model calibration for hydrological prediction in mesoscale Alpine catchments. In the latter, the calibration method uses hydrological process knowledge to extract useful information from very heterogeneous data set available in the region (see also *Schaefli et al., (2005)* and *Schaefli and Huss (2011).*

In other work, *Vidal et al. (2007)* reviewed the process of calibrating physically-based models such as river hydraulic models and distributed hydrological models with a special emphasis on knowledge base calibration. They criticize the fact that calibration is often done without any or with only minimal physical consideration. They advocate a definition of parameter calibration "on the basis of heuristic knowledge gained through modeling experience", and develop a knowledge based calibration support system for hydraulic modelers. The result is an automatic knowledge-based trial and error approach that also has the advantages of reliability and reproducibility. The resulting CaRMA-1 algorithm mimics the way that experts tackle particular calibration cases to obtain the most reasonable calibrated hydraulic model considering the data available. Other examples of limited calibration (parameter adjustment) and hydrologic reasoning for parameters estimation of physically based distributed models can be found in *Feyen et al. (2000)* using MIKE SHE, *Zehe and Bloschl (2004)* for parameter adjustments of CATFLOW, and *Bahremand et al. (2005, 2007), Liu et al. (2003, 2005)* with the WetSpa model, and *Salvadore (2015)* with the WetSpa-Python model.

Some recent publications regarding conceptual hydrologic models have also drawn attention to the use of expert knowledge in parameter estimation and constraining parameter calibration; see for example *Antonetti et al. (2015), Hrachowitz et al. (2014), Gharari et al. (2014), Hellebrand et al. (2011)* and *Viviroli et al. (2009)*. Overall, the examples mentioned above lend support to the author's conviction that by gaining some understanding about hydrologic processes, and by trying to relate the parameters to observable (or conceptual) watershed characteristics, it is possible to infer reasonable values for the parameters of a hydrological model.

In support of this viewpoint, let us look at some examples using the WetSpa model, which has 11 parameters that must be specified (*Liu and De Smedt, 2004*). As a trivial case, consider the parameter Kg_m that represents the maximum active groundwater storage (in mm) and controls the amount of evaporation possible from the water table. This parameter has typically been considered to be "*insensitive*" (see *Bahremand and De Smedt, 2008*), which makes sense of course if the catchment is mountainous and in an upstream area (e.g., catchment order 2), because logic dictates that since the depth to groundwater is so deep, there will be little or no direct evaporation from the water table. In such a case we can save time by fixing this parameter to a large value, and directing our attention to other aspects of the model. Similar reasoning can be applied to several other parameters (*Bahremand et al. 2007, Liu et al. 2003*).

Alternatively, if the practitioner prefers to proceed with an automatic calibration approach (although I prefer the manual calibration approach due to its ability to enhance hydrologic knowledge), much is to be gained by advising her/him to implement some logical relativity restrictions. For example, in the WetSpa model it makes sense to always restrict the value for parameter Kg_i (initial active groundwater storage, in mm) to be less than the value for Kg_m. Doing so helps to restrict the calibration search space, so that the "*best*" parameter values are achieved with the least effort, and the parameter values remain relatively consistent with their conceptual meaning. A nice example of this is provided by *De Smedt et al. (2000)* who implement such reasoning in regards to the parameter values (based on an understanding of the physical structure of the model) and obtain quite good model simulation results without resorting to any "*calibration*". In support of this[AB7], note that *Safari et al (2012)* reported satisfactory results using an uncalibrated WetSpa, with only minor improvements obtained through calibration (see also *Smith et al. 2012*). *Zeinivand and De Smedt (2009, 2010)* reported results of the snow modules of the WetSpa model using preset values with no calibration.

Other "no-calibration" modeling studies using physically-based distributed hydrologic models have reported mixed success (e.g., *Semenova et al. 2015, Venogradov et al. 2011, Refsgaard and Knudsen 1996,* and *Refsgaard et al. 1999)*. Here, "no-calibration" refers to the use of preset parameter values, and "limited-calibration" is taken to mean "manual adjustment … applied to a small group of specially chosen parameters … … carried out as a priori defined narrow ranges of parameter variation…" (*Vinogradov et al. 2011*).

Examples of limited calibration of the WetSpa model are given by *Liu (2003, 2005)* and *Bahremand (2007, 2005)*. I think of such an approach as being a kind of "*white box calibration*", and my experiences with the WetSpa model (*Bahremand et. al 2005 and 2007, Bahremand and De Smedt, 2008 and 2010*) suggest that it can help to ensure a considerable degree of consistency in both the parameter values and the model behavior. As discussed later in this paper, other no-calibration attempts for physical modeling have been reported using the novel approach of optimality (*Schymanski et. al. 2009*), maximum entropy production (*Westhoff and Zehe, 2013*), and behavioral modeling under organizing principles (*Schaefli et al. 2011*).

Of course, when a user selects reasonable initial values for the automated local parameter search, this is akin to bringing some kind of informed prior information to bear on the calibration process, in a manner similar to Bayesian inference, or the expert opinion in decision-making. Accordingly, it helps to improve calibration efficiency, results in enhanced parameter consistency, and reduces uncertainty, thereby improving the overall result. Similarly, in a regionalization process, we bring to bear our prior knowledge about the nature of the catchment and the dominant processes within it to minimize (and if possible, avoid) the need for model calibration and parameter estimation tasks. Via a process of generalization, we find ways to apply our models in ungauged basins based on parameter maps that relate catchment characteristics to parameter values via a combination of expert knowledge and empirical evidence (*Bergstrom, 2006; Bardossy, 2007*). And, in the case of expert opinion used to guide decision-making we employ a similar practice

The point is, that in all of the cases, there is a greater emphasis on process understanding, and as such understanding is enhanced, the parameter estimation problem becomes progressively more trivial. As stated by Hoshin Gupta in a recent email communication (email communication, 31 March 2015), "*it is good to give the students a well-organized frame to think about the model development process because, it can dramatically help to reduce the effort. In my opinion we (the community) have taken a journey of about 30 years long to "rediscover" this because in the late 70's and 80's we were seduced by the ideas of "optimization" (which came from operations research) and the ability to play with computers. Hopefully now the field of "systems hydrology" will focus more on what I like to call the "learning problem" - which is more about architecture and process parameterization than about parameters. Of course some amount of calibration will generally help because the model is always a simplification*".

3 On the Model development process

The model development process follows a series of several steps. Since these steps have been discussed variously by *Beven (2012)*, *Gupta et al. (2012)*, and *Gupta and Nearing (2014),* among others, the reader may refer to those articles for details. I mention them only briefly here. As mentioned by *Gupta et al. (2012)* first stage is informal and involves the formation of "*perceptions*" about the system. In the formal steps, we begin with a "*conceptual model*", and then proceed (in the language of Beven) to develop a "*procedural model*" (but see *Gupta et al., 2012* for considerably more fine-grained detail). Finally we run the model with some initial parameter guesses, and then proceed with model calibration and evaluation, sensitivity analysis and uncertainty analysis. These last 4 steps can perhaps be grouped under the general term of "*model optimization*".

The important step that follows is that of model "*verification*" (or perhaps we can call this diagnostic evaluation and improvement; see *Gupta et al., 2008*). In *Beven (2012)* is implied by the word "*revise*" (in the second illustration of the first chapter of Beven's book). We advise the practitioner that if the constructed model "*fails*" the diagnostic evaluation step we should first revisit the calibration step (just one step back) to check whether we could do better by calibrating our model differently. If everything is found to be "ok" in this step, we should proceed backward one more step and take a closer look at the "*procedural model*", to check the computer code for errors. And, if this seems fine we can proceed to examine our "*conceptual model*", whereby we check the equations used, the manner in which subsystems are linked to each other, inputs, outputs, functions, and so on. Finally if everything seems fine, then we may be forced to question our perceptions, examining in detail how we have defined the processes.

However, the current modeling practice seems to be largely stuck in the model optimization stages. *Gupta and Nearing (2014)* correctly suggest that we have given more than enough attention to the problem of model optimization. And several authors have argued that if we want to have real improvements in modeling practice and performance, then we need to take a more serious look at the early steps in the modeling protocol, and in particular focus in on the "*process model*" (even being willing to alter our perceptual model).

It is instructive to note that, despite the diversity in hydrological behaviors found in catchments of different kinds, most current conceptual watershed models are only slightly different implementations of very similar perceptions and conceptions in regard to watershed behavior, and involve very similar kinds of simplifications and assumptions. In this context, novel ideas such as HAND and the topographic index embody interesting revisions in the perceptual and conceptual model stages of conceptual-hydrologic modeling (*Savenije, 2010; Gharari et al., 2011; Gao et al., 2014*). Similarly the REW approach is an example of revisions in early stages of physical-hydrologic modeling (*Reggiani et al*[AB8]*. 1998 and 1999)*. And as suggested by *McDonnell et al. (2007)*, "*New approaches should rely not on calibration, but rather on systematic learning from observed data, and on increased understanding and search for new hydrologic theories*". It is, of course always easier to improve upon an already existing model/framework. In some cases, however, really significant improvements can only come about by starting at the very beginning. In my view, the end of optimization can serve as a new beginning for the hydrological modeling process.

4 On the modeling and evaluation of hydrologic processes

It seems obvious that hydrologists should be ready to investigate our perceptions and be willing to make dramatic improvements in conceptualizations as needed. Various assumptions, expediencies and simplifications may need to be changed or disregarded. As mentioned by Grey Nearing in a recent email communication (email communication, 31 March 2015), "*It is strange that we know a priori that any model we build will be incorrect, and so the pertinent question in my mind is in what sense a wrong model can be useful. Since calibration can never fix the fact that our models are always wrong, we must interpret the calibration procedure as in some sense reducing the impact of our model's errors on the utility of that model. Neither calibration nor iterative model refinement will ever result in a correct model, and error functions, likelihoods, objective functions, and performance metrics are all attempts to measure model utility, not model correctness. My opinion is that this utility approach to model building and model evaluation is misguided. Instead of building a model that we know is wrong and then trying to estimate how wrong it is, we should try to use our knowledge of physics to constrain the possibilities of future events. That is, instead of trying to approximately solve complex systems of equations, use the equations to limit the possibilities of future events. Shervan Gharari takes this perspective to assigning parameters in his recent paper (Gharari et al., 2014), and for this reason it is one of my favorite*".

While Nearing argues that the *current* paradigm is based fundamentally around a concept of utility, and that our knowledge of physics should be used to constrain the possibilities of future events, Gupta refers to such a focus as "*prediction and problem solving, and to serve such purpose while improving our understanding of "physics", so the target becomes the "model" and this sets up a recursive loop when we try to "support/evaluate" the model*."

In practice, I have found a ladder type (tree-like) evaluation and model intercomparison framework (of flexible length) to be useful for model evaluation. In the short version of this ladder, the modeler is able to "*evaluate/support*" a particular model by seeking, for example, an improved simulation of the total hydrograph. Given a lumped conceptual model "A" and a physics based distributed model "B", the short ladder evaluation allows us to compare the hydrographs simulated by A and B with each other, and with the observed target data. This kind of evaluation really just serves the model, in the sense that it supports the specific kind of prediction needed by a target application such as river hydrograph simulation/prediction.

In contrast, the long version of the ladder can take us much deeper. In this type of evaluation, our goal is not model intercomparison based on target performance, but is instead based on consistency or realism. For example, in the first step (stair/stage) we have a descriptive table that enables comparison between the conceptualizations underlying the models. It enables us to compare which hydrological processes are represented in the models, and how they are interlinked (although this latter could perhaps be considered a second step). In such a context, it does not really make sense to compare an artificial neural network black box type model against a fully distributed physically-based model, which comparison could mislead a naïve practitioner (being a comparison between two different kinds of things).

Ultimately, we need to develop frameworks for model evaluation and comparison that enable us to give more weight to ones that better represent the underlying physics (see *Clark et al., 2011; 2015a,b; Mendoza et al., 2015*). This kind of long ladder evaluation enables us to progressively deepen our understanding, step by step. Along the way, some models may be left behind, but can continue to serve our immediate and intermediate needs such as for hydrograph simulation. However, later steps may require our model to pass additional tests, such as requiring the flow velocity in streams of order 1 and located in forested terrain to be meaningful in comparison with the velocities in similar streams passing through high altitude farmland.

In such a context, a simple hydrograph comparison may generally not be sufficient, and simple model efficiency and performance metrics on streamflow will not guarantee that the system has been correctly described (*Klemes, 1986; Bergestrom, 1991;* see also *Savenije, 2009* for a discussion of what constitutes a "*good model*"). So, for example, the behavioral and non-behavioral models partitioning within a GLUE framework (*Beven and Binley, 1992*) should not be based simply on model output-based performance criteria, but should be meaningful and correct in an intellectual manner. The use of relational rules (as in *Gharari et al., 2014*) serves the function of prior information. {here I deleted 6 lines}[AB9]

As has been pointed out in the literature, our approach to model evaluation that is based in performance criteria also needs improvement. Recent work in this regard includes the Kling-Gupta efficiency (*Gupta et al., 2009*), the increasing emphasis on process/signature-based diagnostics (*Gupta et al., 2008; Yilmaz et al., 2008*), and the use of multi objective criteria and evaluation on multiple variables (*Gupta et al., 1998; Pechlivanidis and Arheimer, 2015*). Equally important, we need to establish benchmark problems that serve as a set of standard test cases, thereby providing the modeling community with a way to perform fair assessments of competing formulations, parameterizations and algorithms (*Maxwell et al., 2014; Paniconi and Putti, 2015*).

Ultimately, model optimization can help establish the best possible model performance compared with input-output data, uncertainty analyses can help to reveal model structural deficiencies, and comparison against benchmark prediction limits (e.g., *Schaefli and Gupta 2007*) can provide a possible way of checking the correctness of our understanding of the hydrological processes at a given time and place (*Montanari and Koutsoyiannis, 2012*). While this may be obvious to an experienced modeler, I feel that we should be thinking about building a structured framework that can help beginners/students to stay on the right track, and not be deceived by "good" values of summary metrics such as the Nash-Sutcliffe Efficiency. In such a structured framework, it will be important to take first into account model simplifications, assumptions, formulations, the code, and the list of processes, before examining the simulation results. And, an automated model calibration procedure should not be used as a way to justify a poorly formulated model that is then "*camouflaged by uncertainty estimation*". As has been pointed out before many times (see e.g., *Semenova and Beven, 2015*), expert opinion and judgment should matter when evaluating the credibility of model performance and predictions. To this one might add that scientific knowledge and principles of physics should matter even more, as should practical perceptual and observational knowledge about the system being modeled.

As examples of the latter, consider the following. Although flow widths change along the stream network, most hydrological models use a constant width for the stream network; at the very least, streams of different order should be allocated different widths. Most hydrological models assume constant flow velocity fields for the entire duration of the simulation; in fact, flow velocities should be considered together with the sediment and bed loads. Similarly, hydrological flow routing should take into account transmission losses, the differences between velocities and celerities, hysteresis with respect to total storage in a landscape element, heterogeneities and the extremes of their distribution. To quote *Semenova and Beven (2015)*, "*These are requirements for any distributed modeling scheme in hydrology that is going to be intellectually satisfying in reproducing both flow and travel times of water*". Doing so will bring to bear well-known hydraulic principles. Bringing physics and more detailed attention to process modeling will also leads to better integration of surface and subsurface hydrology in models (*Paniconi and Putti 2015*).

Moreover, alternative theories and approaches, such as representative elementary watershed concept of *Reggiani et al. (1998 and 1999)* and the thermodynamic reinterpretation of the HRU concept of *Zehe et al. (2014)*, help us to limit uncertainty and better deal with equifinality by improving our understanding of the system. Although even physics based models face equifinality (see *Klaus and Zehe, 2010; Weienhoefer and Zehe, 2014*), as this problem simply arises from the structure of our equations (see *Zehe et al., 2014*), by explicitly disentangling driving gradients and resistance terms in flow equations the process-based models offer more options to exert constraining rules to end up with a rather unique parameter set (*Zehe et al., 2014*). Taking more processes into account decreases non-uniqueness, as for example *Wienhöfer and Zehe (2014)* reduced "the number of equifinal model set-ups" by the results of solute transport simulations.

Also, some processes such as subsurface processes and preferential flow need to be better represented explicitly, and we should consider the limitation of Darcy-Richards equations (being diffusive and assuming local equilibrium conditions) regarding the fast advective responses and cell size limitation (*Vogel and Ippisch, 2008*). Similar to the multi-objective criteria approach in model optimization, where a set of criteria is involved in the search for a unique parameter set; accordingly from a different angle, if we take more physical processes into account into our model structure, it does a similar thing, i.e. it gives us more options to constrain parameter values and reach a rather unique parameter set. Therefore, the equifinality should be dealt with from different angles to help us to arrive at a better model.

Another approach to dealing with equifinality[AB10] is by limiting the parameter values through a procedure that can be called parameter allocation. In the following section, I express my ideas in this regard and on the future of hydrological modeling.

5 On parameter allocation and the future of hydrological modelling[AB11]

In this section, I articulate my opinions regarding parameter allocation and the future of hydrological modeling, and in particular my opinion in regards to physically-based distributed models as the right path to model hydrologic processes and to avoid calibration and its related uncertainties.

5.1 Contrasting parameter calibration and parameter allocation

In the process of model development, calibration seems unavoidable (*Beven, 2001; Montanari and Toth, 2007; Hrachowitz et al. 2013*) as a way to compensate for our lack of knowledge of spatial heterogeneities in watershed properties and our lack of understanding of hydrologic processes (*McDonnell et al. 2007*). It can be done either manually or automatically or by some hybrid approach (*Boyle et al 2000, Hogue et al 2000, 2006*). Manual calibration applies hydrologic knowledge and reasoning to obtain the good parameter values in fewer attempts but involves trial and error and is very time consuming. Automated calibration approaches may not add much to the hydrologic knowledge of the practitioner, but can be very helpful when there are many parameters to be determined (overcoming the tedium and time involved in manual calibration), provides the possibility of quickly checking numerous combinations of plausible parameter values (that would be impossible to attempt manually), and can provide useful support to model diagnostic evaluation. Indeed, when the best parameter estimate is physically unrealistic, one may conclude that the model is not adequate, and such a conclusion can only be reached if an exhaustive search for the best parameter estimates has been carried out (see Montanari's referee comment on this paper; *Gupta et al. 1999*). Since, automatic calibration is an iterative procedure, it also provides information useful for parameter sensitivity and uncertainty analysis *(Bahremand and De Smedt, 2008)*. As explored by *Boyle et al (2000)* and *Hogue et al (2000, 2006)*, a hybrid combination of these two types of calibration approaches is also possible.

Meanwhile, what I refer to here as parameter "allocation" does indeed play an important role in hydrological modeling but has not received sufficient discussion although it is something that experienced modelers typically do in any modeling study (see Schaefli's referee comment on this paper). I argue that this aspect deserves more attention, since it is in the direction of achieving more understanding of the hydrological processes, the way they are represented in the model, and the link between model parameters and catchment characteristics (this understanding can be extended to conform with the organizing principles mentioned in *Schaefli et al., 2011*).

Parameter allocation is relevant in the case of process-based models, whose parameters are more likely to have physical or conceptual meaning and be rationally explainable. With some degree of practice, and after having gained some understanding of how hydrological processes are represented in the model and how the parameters relate to observable or conceptual catchment characteristics, the modeler can specify values for the parameters based on logical reasoning. Of course, for some of the parameters, a few trial and error adjustments might still prove to be necessary and useful. It is, therefore, a heuristic technique, a kind of ansatz, in which an educated guess is made regarding the parameter values, which can later be verified through an evaluation of the model performance.

So, parameter allocation can be viewed as a part of (or kind of) the parameter calibration procedure. Whether using a manual or automatic approach, the modeler can use rationality and logic (based mainly on hydrologic reasoning) to guide parameter improvements. Reasoning can be used to establish constraints and relational rules between parameters, in accord with relevant organizing principles (this needs to be elaborated via future modeling research), and in accordance with a higher level (global or regional) water balance model. These latter two (conformity with organizing principles and water balance scheme) are particularly relevant when attempting to develop a community hydrological model (*Weiler and Beven, 2015*) or a hyper resolution model of everywhere (*Beven, 2007, 2015, Beven and Alcock, 2012*). Such constraints and relational rules can either be applied manually, or by some computer-based procedure (see *Gharari et al. 2014; Vidal et al. 2007*).

Essentially, what makes the difference between parameter “allocation” and parameter “calibration” is the extent of prior knowledge applied by the modeler. In parameter calibration, prior knowledge is mainly used to set the allowable range of parameter values (to establish the “feasible” parameter space). In parameter allocation, additional prior knowledge is imposed in the form of relational rules between parameters, some certain constraints and principles. In this case, the modeler does attempts to allocate values for as many of the parameters as possible, so that the need for trial and error adjustments is minimized and limited to only a few parameters.

The point is, of course, to make as much use of prior knowledge as possible, so as to limit/minimize the uncertainty, while arriving at reasonable (physically or conceptually defensible) values for the parameter, ones that support our basic conceptual understanding of the system. In this context, models with the smallest number of “parameters-subjected-to-calibration” will be considered more scientifically interesting, and parameter estimation becomes part of the learning process (see comment by Hoshin Gupta mentioned above). The primary motivation and emphasis becomes “understanding” rather than “good results”; i.e., less accurate results with reasonable parameter values (and model behaviors) are more desirable than more accurate results with unreasonable parameter values. It brings to the foreground the need to make a tradeoff between accuracy and reasonability, given the fact that every model is a simplification of reality.

Below, I outline a few steps that can be followed in the parameter allocation procedure for a physics based model:

I) Conduct a preliminary rough evaluation of parameter behavior or sensitivity (an optimum parameter set from a previous study in a different catchment can be a good choice to start with). The modeler is supposed to understand how the model response relates to the values of its parameters, and such a test helps to verify the expected behavior for the new study area.

II) Specify (allocate) values for those parameters for which approximate values can be easily established by following rules of thumb (like parameters Kg_i and Kp in the WetSpa model, see *Bahremand and De Smedt, 2008* for the model parameters).

III) Fix any “insensitive” parameters to reasonable nominal values. This step may not generally be necessary for physically-based distributed models, because their parameters are usually likely to be sensitive; however, in my work with the WetSpa model, I found it appropriate to fix one insensitive parameter (parameter Kg_m). Similarly *Roux et al. (2011)* and *He et al. (2015)* also report fixing insensitive parameters of their physically based models (MARINE and THREW).

IV) Allocate approximate values for parameters that show consistent relational behavior with catchment characteristics (e.g., parameter Kg in the WetSpa model, see *Bahremand et al. 2005, 2007*, *Liu et al. 2003, 2005*).

V) Collect and list all of the relational inequality constraints between parameters (e.g, $Kg_i < Kg_m$ in the WetSpa model), the conceptual relations between parameters and catchment characteristics, (as well as organizing principles and water balance related constraints).

VI) Apply inequality conditions that may be relevant between some of the parameters. Those parameters having constraints and relational rules are allocated together. The constraints can be either implemented manually or using simple computer codes in case of automatic procedure (see, for example, the tool presented by *Vidal, 2007*).

VII) In some cases, the model parameters and/or processes will be required to conform with organizing principles such as optimality, landscape evolution laws, and Horton laws of stream networks (e.g. Horton number of bifurcation); and a higher level water balance model (a regional or global model) should be satisfied. As an example of the latter, *Schaefli and Huss (2011)* used glacier mass balance data to constrain the parameter uncertainty for their hydrological model in a glaciered basin (see also *He et al. 2015*). For the purpose of developing a community hydrological model, a universal water balance model can be used to establish constraints on our local model and its parameters. Another way to say this is that while our models are calibrated locally to observations, they must also obey parameter inter-relationships and constraints, and the organizing principles and components of a universal water balance model. These three different types of constraints (i.e, constraints between parameters, organizing principles, and balance related controls) will allow us to pre-set most of the parameters. However, the idea behind this step still feels somewhat "rough" in my minds, and needs further elaboration and perhaps revision.

As mentioned above, for some of the parameters the results will be a parameter range rather than a definite value, and it is likely that some residual manual trial and error adjustments may still be necessary before the modeler can decide on the final parameter values. Having arrived at this "allocated" set, one must trust in, and be confident with, the outcome.

5.2 Some further comments regarding parameter allocation

My experience with this kind of parameter allocation is that it has attributes of both the bottom-up and top-down approaches to model development. By this, I mean that the modeler is required to be able to change her/his viewpoint based on what happens during the parameter allocation process. The manual-expert and automated approaches each have their advantages and disadvantages, and an experienced modeler brings both approaches to bear when seeking to allocate values for the parameters. In this way, the process can act as a link between deductive physics-based distributed modeling and the behavioral modeling approach (using organizing principle to constrain models) described by *Schaefli et al. (2011)*.

Whereas parameter allocation can be used to establish relatively narrow ranges on the parameter values, the application of optimality or organizing principles can help to further restrict these ranges. *Schaefli et al., (2011)* express this as "*adjusting the model structure and parameters so as to respect this organizing principle*". Some that have received attention in the literature include the optimality principle (*Schymanski, 2008 and 2009*), maximum energy dissipation (*Zehe et al. 2010*), maximum entropy production (*Kleidon and Schymanski, 2008; Kleidon et al. 2012 and 2013; Westhoff and Zehe, 2013*), landscape evolution laws and optimal channel networks (*Rodriguez-Iturbe and Rinaldo, 2001; Rinaldo et al. 2013*) or self-organized dissipation of singular events (*Beven 2015).* Proper application of such principles can be used to improve the theoretical underpinnings of hydrologic models *(Clark et al 2016)* and can provide constraints that might be useful in making predictions *(Schaefli et al. 2011)*; although see *Beven (2015)* who calls them purely theoretical conjectures that are difficult to prove. *Schymanski et al. (2009)* presents a good example of how optimality may be a useful way of approaching the prediction and estimation of some vegetation characteristics and fluxes in ungauged basins without calibration.

5.3 On the future of hydrological modeling

To reiterate, hydrological modeling has become more and more physics- and process-based. This opinion paper reflects my passion for process-based models, and my (perhaps) radical belief that other types of models do not serve us well anymore. When working with process models, we should spend less time on model optimization and instead focus on our perceptual and conceptual insights with a view to better understanding and expressing the physical nature of the system. This implies that: [AB12]

1) models should typically only contain physically based parameters
2) models having fitting parameters without physical basis are inferior and should be abandoned
3) spatially-lumped parameters are not physically based and should be avoided
4) models with physically based parameters that are unable to reproduce observations are incomplete or erroneous and need to be improved, fixed or abandoned

5) models with non-sensitive parameters are basically inadequate to simulate the system (i.e., over-parameterization is bad)

6) physical models that "fail" need to be improved, and can help us learn something about what is wrong (impetus for research)

7) in the limit we should strive for "white box models" that do not need any calibration, or only minor calibration (parameter adjustment).

To reach such a goal we need to apply better measurements[AB13] and better physics. As stated by *Paniconi and Putti (2015)*, "*no one would disagree that scientific progress requires a constant dialogue between measurement, analysis, and simulation*". The *Gupta et al. (2014)* paper advocating large-sample hydrology also implies the necessity of such dialog to improve hydrologic science, and *Hrachowitz et al. (2013)* mentions "data" as the backbone of any type of progress.

Of course, both involve significant challenges. *Beven and Germann (2013)* provide a thoughtful discussion on the misuse of physics in simulating flow through porous media, and in particular, the limitations of Darcy and Richards equations; they suggest the representation of preferential flows via a Stokes flow for profile scale and multiple interacting pathways model *(Davies et al., 2011)* at the hillslope scale. *Zehe et al. (2013)* propose a thermodynamic approach to represent catchment scale preferential flow. The mass, energy and momentum balance closure problem presents a significant challenge (*Beven, 2006a*, see also the editor's comment on my paper), although there has been some progress (*Reggiani et al. 2000, Reggiani and Schellekens, 2003, Reggiani and Reintjes, 2005, Tian et al, 2006, Mou et al. 2008*). *Kleidon and Schymanski (2008*) suggest that the optimality principle can help with the scaling of hydrologic fluxes; knowing the hydrologic fluxes at a larger scale can provide a "big" picture, and a top-down approach can be used to infer the boundary fluxes of ungauged basins at smaller scales.

Perhaps we can describe the future of hydrological modeling by means of an analogy with the problem of solving a spherical jigsaw puzzle, where the puzzle involves assembly of numerous oddly shaped interlocking and tessellating pieces, each having only a small part of the overall picture. To solve the puzzle it is helpful to have 4 different kinds of information:

1) A sense of the complete picture; this can be compared with our perceptual and conceptual model of the hydrologic cycle at the global scale.

2) Information regarding the puzzle edges (borders); this is analogous with large-scale water balance and its components

3) Information regarding the picture expressed by each piece itself; this is analogous to regional or catchment scale hydrological models (the representation of local scale hydrological processes)

4) Information regarding the ways in which the pieces interlock.

It is well known that rapid solution of a jigsaw puzzle can be facilitated by sorting and categorizing the pieces according to shape, color, edge and corner shapes, and shapes of interlocking connectors; this may be comparable with concepts such as generalization, regionalization, and the organizing principles and behavioral modeling of *Schaefli et al. (2011)*. Comparing the partially constructed puzzle with the complete picture (usually printed on the front of the box) is similar to what I have described as a mind commute between the top-down and bottom-up viewpoints (Sivapalan, 2005). The learning process emphasized by *Beven (2007)* in his "*models of everywhere*" and the "*learning instead of rejection*" view exposed by *Gupta and Nearing (2014)* is expressive of this practice. As we continue to work on the puzzle, we try to build upon already completed sections, and eventually we get to the stage where we can see the end of the project where the "holes" become the objects of our attention.

6 Conclusions

In conclusion, it is clear that we need to make a determined effort to shift the focus of our modeling studies away from parameter optimization and towards a deeper attention to process modeling and revision of our conceptual models. We should even be ready to revise our perceptual models. *Gupta and Nearing (2014)* argue that we need robust and rigorous methods to support such a shift, and *Gharari et al. (2014)* shows that such an approach can help to liberate us from the need for model calibration, transforming it into a process of parameter allocation. Ideally, the calibration and evaluation procedures would act synergistically to drive model improvement. Hopefully then, we will move past "*equifinality*" to achieve "equimodellity", reaching at last one fulfilling model that is a "*model that is so physically correct that it does not need calibration at all*"(the third aforementioned solution of Bergstrom). Although such a target might seem unreachable, it could at least act as a beacon for hydrologists.

Acknowledgements

I would like to thank Hoshin Gupta for his constructive comments and editing the manuscript, and for encouraging me to write and submit my opinion. The paper was significantly improved after being refined by Hoshin Gupta, twice (the first and the second versions were both trimmed and enhanced by him, so I really owe Hoshin a lot for his invaluable help). Prof. Florimond De Smedt my PhD promoter helped me during the review process (e.g. the first paragraph of subsection 5.3), I really appreciate his scientific support and valuable advices. I would like to thank very much the referees, i.e., Keith Beven, Alberto Montanari and Bettina Schaefli, and the editor Erwin Zehe for their constructive review comments. The paper improved very much according their very useful comments, questions, instructions and supports. I would also like to thank Grey Nearing, Shervan Gharari, Claudio Paniconi, Ali Safari, Yongbo Liu, H.H.G Savenije, Hamidreza Sadeghi, Vahedberdi Sheikh, Hossein Zeinivand and Arashk Holisaz for their useful comments on the manuscript. I also thank Massimiliano Zappa, Thomas Bosshard, Olga Semenova, Lyudmila Lebedeva, Mohsen Tavakoli, Jan Corluy and Stanislaus Shymanski for encouraging emails and sending me their papers. I appreciate some English corrections done by Julie Deconinck on the very first version of the manuscript. I thank the journal authorities for waiving the article processing charges, it is very much appreciated.

References

Antonetti, M., Buss, R., Scherrer, S., Margreth, M., and Zappa, M.: Mapping dominant runoff processes: an evaluation of different approaches using similarity measures and synthetic runoff simulations, Hydrol. Earth Syst. Sci. Discuss., 12, 13257-13299, doi: 10.5194/hessd-12-13257-2015, 2015.

Bahremand A, Corluy J, Liu YB, De Smedt F, Poorova J, and Velcicka L.: Stream flow simulation by WetSpa model in Hornad river basin, Slovakia. In: van Alphen J, van Beek E, Taal

M (eds) Floods, from defence to management. Taylor–Francis Group, ISBN 0 415 38050 2, London, pp 67–74, 2005.

Bahremand A, De Smedt F, Corluy J, Liu Y. B, Poórová J, Velcická L, and Kuniková E.: WetSpa model application for assessing reforestation impacts on floods by in Margecany–Hornad watershed, Slovakia. Water Resour Manag 21:1373–1391, doi: 10.1007/s11269-006-9089-0, 2007.

Bahremand, A. and De Smedt, F.: Distributed Hydrological Modeling and Sensitivity Analysis in Torysa Watershed, Slovakia, Water Resources Management, 22:393-408, doi: 10.1007/s11269-007-9168-x, 2008.

Bahremand, A. and De Smedt, F.: Predictive Analysis and Simulation Uncertainty of a Distributed Hydrological Model, Water Resources Management, 24(12), 2869-2880, doi: 10.1007/s11269-010-9584-1, 2010.

Bardossy, A.: Calibration of hydrological model parameters for ungauged catchments, Hydrol. Earth Syst. Sci., 11(2), 703–710, 2007.

Bergstrom, S.: Principles and Confidence in Hydrological Modelling, Nordic Hydrology, 22, 1991, 123-136, 1991.

Bergstrom, S.: Applications of the HBV hydrological model in prediction in ungauged basins. In: Large Sample Basin Experiments for Hydrological Model Parameterization Results of the Model Parameter Experiment MOPEX. IAHS Publ. 307. 97-107, 2006.

Beven, K. and Binley, A.: THE FUTURE OF DISTRIBUTED MODELS: MODEL CALIBRATION AND UNCERTAINTY PREDICTION, Hydrol. Process., 6, 279–298, 1992.

Beven, K.: Changing ideas in hydrology - The case of physically-based models. Journal of Hydrology, 105(1-2), 157–172, doi: 10.1016/0022-1694(89)90101-7, 1989.

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.

Beven, K.: How far can we go in distributed hydrological modelling? Hydrology and Earth System Sciences, 5(1), 1–12, doi: 10.5194/hess-5-1-2001, 2001.

Beven, K.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, Hydrol. Process., 16: 189–206. doi: 10.1002/hyp.343, 2002.

Beven, K. J.: Searching for the Holy Grail of scientific hydrology: Q_t=H (S←,R ← ,Δt) A as closure, Hydrol. Earth Syst. Sci., 10, 609–618, 2006a.

Beven, K.: A manifesto for the equifinality thesis. Journal of Hydrology, 320(1-2), 18–36, doi: 10.1016/j.jhydrol.2005.07.007, 2006b.

Beven, K.: Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. Hydrol. Earth Syst. Sci, 11(1), 460–467. Retrieved from www.hydrol-earth-syst-sci.net/11/460/2007, 2007.

Beven, K. J., Smith, P. J., and Wood, A.: 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, 2011.

Beven, K. J.: Rainfall-Runoff modeling: The primer, Second edition, John Wiley and Sons, Chichester, UK, 2012.

Beven, K. J. and Alcock, R. E.: Modelling everything everywhere: A new approach to decision-making for water management under uncertainty, Freshwater Biology, doi: 10.1111/j.1365-2427.2011.02592.x, 2012.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, doi: 10.1002/wrcr.20156, 2013.

Beven, K. and Smith, P.: Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models, Journal of Hydrologic Engineering, 4014010(15), A4014010, doi: 10.1061/(ASCE)HE.1943-5584.0000991, 2014.

Beven, K.: What we see now: Event-persistence and the predictability of hydro-eco-geomorphological systems, Ecological Modelling, 298, 4–15, doi: 10.1016/j.ecolmodel.2014.07.019, 2015.

Beven, K., Cloke, H., Pappenberger, F., Lamb, R., and Hunter, N.: Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface. Science China: Earth Sciences, 58: 25–35, doi: 10.1007/s11430-014-5003-4, 2015.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Towards Improved Calibration of Hydrologic Models: Combining the Strengths of Manual and Automatic Methods, Water Resources Research, Vol. 36, No.12, pp. 3663-3674, doi: 10.1029/2000WR900207, 2000.

Blöschl, G., Reszler, C., and Komma, J.: A spatially distributed flash flood forecasting model, doi: 10.1016/j.envsoft.2007.06.010, 2008.

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47, W09301, doi: 10.1029/2010WR009827, 2011.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, 2498-2514, doi: 10.1002/2015wr017198, 2015a.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., and Marks, D. G.: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies, Water Resour. Res., 51, 2515-2542, doi: 10.1002/2015wr017200, 2015b.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Jackson, B. M., Freer, J. E., Arnold, J. R, Moore, R. D., Istanbulluoglu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models. Water Resour. Res., 52, doi: 10.1002/2015WR017910, 2016.

Davies, J., Beven, K., J., Nyberg, L., and Rodhe, A.: A discrete particle representation of hillslope hydrology: Hypothesis testing in reproducing a tracer experiment at Gårdsjön, Sweden, Hydrol. Process., 25, 3602–3612, doi: 10.1002/hyp.8085, 2011.

De Smedt, F., Liu, Y. B., and Gebremeskel, S.: Hydrological modeling on a catchment scale using GIS and remote sensed land use information, in: Brebbia CA (ed) WTI, Boston, pp 295–304, 2000.

Feyen, L., Vázquez, R., Christiaens, K., Sels, O., and Feyen, J.: Application of a distributed physically-based hydrological model to a medium size catchment, Hydrol. Earth Syst. Sci., 4, 47-63, doi:10.5194/hess-4-47-2000, 2000.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested of the Upper Heihe, China, Hydrol. Earth Syst. Sci., 18, 1895-1915, doi: 10.5194/hess-18-1895-2014, 2014.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, Hydrol. Earth Syst. Sci., 15, 3275-3291, doi: 10.5194/hess-15-3275-2011, 2011.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18, 4839–4859, doi: 10.5194/hess-18-4839-2014, 2014.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resour. Res., 34(4), 751–763, doi: 10.1029/97WR03495, 1998.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Status of Automatic Calibration for Hydrologic Models: Comparison With Multilevel Expert Calibration, doi: 10.1061/(ASCE)1084-0699(1999)4:2(135), 1999.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, Volume 22, Issue 18, pages 3802–3813, doi: 10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, Volume 377, Issues 1-2, Pages 80-91. doi: 10.1016/j.jhydrol.2009.08.003. ISSN 0022-1694, 2009.

Gupta, H. V., Clark, M. P., Vruft, 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.

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

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18, 463-477, doi: 10.5194/hess-18-463-2014, 2014.

He, Z. H., Tian, F. Q., Gupta, H. V., Hu, H. C., and Hu, H. P.: Diagnostic calibration of a hydrological model in a mountain area by hydrograph partitioning, Hydrol. Earth Syst. Sci., doi: 10.5194/hess-19-1807-2015, 2015.

Hellebrand, H., Müller, C., Matgen, P., Fenicia, F., and Savenije, H.: A process proof test for model concepts: modelling the meso-scale, Phys. Chem. Earth, 36, 42–53, doi:10.1016/j.pce.2010.07.019, 2011.

Hingray, B., Schaefli, B., Mezghani, A., and Hamdi, Y.: Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments. Hydrol. Sci. J. 55(6), 1002–1016, 2010.

Hogue, T. S., Sorooshian, S., Gupta, H. V., Holz, A., and Braatz, D.: A Multi-Step Automatic Calibration Scheme (MACS) for River Forecasting Models, Journal of Hydrometeorology, Vol. 1, No. 6, pp. 524-542, doi: 10.1175/1525-7541(2000)001<0524:AMACSF>2.0.CO;2, 2000.

Hogue, T., S., Gupta, H. V., and Sorooshian, S.: A 'User-Friendly' Approach to Parameter Estimation in Hydrologic Models, MOPEX special issue in Journal of Hydrology, 320, pp. 202-217, doi: 10.1016/j.jhydrol.2005.07.009, 2006.

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.

Hrachowitz, M., Fovet, O., Ruiz, L.,Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resour. Res., 50, 7445–7469, doi: 10.1002/2014WR015484, 2014.

Klaus, J., and Zehe, E.: Modelling rapid flow response of a tile-drained field site using a 2d physically based model: Assessment of 'equifinal' model setups, Hydrological Processes, 24, 1595-1609, 10.1002/hyp.7687, 2010.

Kleidon, A., and Schymanski, S.: Thermodynamics and optimality of the water budget on land: A review. Geophysical Research Letters, doi: 10.1029/2008GL035393, 2008.

Kleidon, A., Zehe, E., and Lin, H.: Thermodynamic limits of the critical zone and their relevance to hydropedology, In: Lin, H. (Ed.), Hydropedology, Elsevier, NewYork, pp. 243–281, 2012.

Kleidon, A., Zehe, E., Ehret, U., and Scherer, U.: Thermodynamics, maximum power, and the dynamics of preferential river flow structures at the continental scale. Hydrology and Earth System Sciences, doi: 10.5194/hess-17-225-2013, 2013.

Lebedeva L., Semenova O., and Vinogradova T.: Simulation of Active Layer Dynamics, Upper Kolyma, Russia, using the Hydrograph Hydrological Model, Permafrost and Periglac. Process., 25, pages 270–280, doi: 10.1002/ppp.1821, 2014.

Liu, Y.B., Gebremeskel, S., De Smedt, F., Hoffmann, L., Pfister, L.: A diffusive transport approach for flow routing in GIS-based flood modeling, Journal of Hydrology, 283(1-4): 91–106, 2003.

Liu, Y. B., and De Smedt, F.: WetSpa extension, documentation and user manual, Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels, Belgium, 2004.

Liu, Y. B. and De Smedt, F.: Flood Modeling for Complex Terrain Using GIS and Remote Sensed Information. Water Resources Management, 19(5), 605–624, doi: 10.1007/s11269-005-6808-x, 2005.

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J., Ferguson, I. M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S. J., Kumar, M., Lopez, S., Niu, J., Paniconi, C., Park, Y., Phanikumar, M. S., Shen, C., Sudicky, E. A., and Sulis, M.: Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resour. Res., 50, 1531–1549, doi: 10.1002/2013WR013725, 2014.

McDonnell, J. J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M.L., Selker, J., and Weiler, M.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resour. Res., 43, W07301, doi: 10.1029/2006WR005467, 2007.

McDonnell, J. J. and Beven, K.: Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities, and residence time distributions of the headwater hydrograph, WaterResour.Res., 50, 5342–5350, doi:10.1002/2013WR015141, 2014.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., and Gupta, H. V.: Are we unnecessarily constraining the agility of complex process-based models? Water Resour. Res., 51, 716-728, doi: 10.1002/2014wr015820, 2015.

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.

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

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., Bahr, J., Blöschl, G., Cai, X., Mackay, D. S., Michalak, A. M., Rajaram, H., and Sander, G.: Fifty years of Water Resources Research: Legacy and perspectives for the science of hydrology, Water Resour. Res., 51, 6797-6803, doi: 10.1002/2015WR017998, 2015.

Mou, L., Tian, F., Hu, H., and Sivapalan, M.: Extension of the Representative Elementary Watershed approach for cold regions: constitutive relationships and an application, Hydrol. Earth Syst. Sci., 12, 565-585, doi: 10.5194/hess-12-565-2008, 2008.

Paniconi, C. and Putti, M.: Physically based modeling in catchment hydrology at 50: Survey and outlook, Water Resour. Res., 51, 7090-7129, doi: 10.1002/2015WR017780, 2015.

Pechlivanidis, I.G. and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case, Hydrol. Earth Syst. Sci. Discuss., 12, 2885-2944, doi: 10.5194/hessd-12-2885-2015, 2015.

Ragettli, S. and Pellicciotti, F.: Calibration of a physically based, spatially distributed hydrological model in a glacierized basin: On the use of knowledge from glaciometeorological processes to constrain model parameters, Water Resour. Res., 48, W03509, doi: 10.1029/2011WR010559, 2012.

Refsgaard, J. C., and Knudsen, J.: Operational Validation and Intercomparison of Different Types of Hydrological Models, Water Resour. Res., 32(7), 2189–2202, doi: 10.1029/96WR00896, 1996.

Refsgaard, J. C., Thorsen, M., Jensen, J. B., Kleeschulte, S. and Hansen, S.: Large scale modelling of groundwater contamination from nitrogen leaching, J. Hydrol., 221(3–4), 117–140, doi: 10.1016/S0022-1694(99)00081-5, 1999.

Reggiani, P., Hassanizadeh, S. M., and Sivapalan, M.: A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, Adv. Water Resour., 22, 367–398, 1998.

Reggiani, P., Hassanizadeh, S. M., Sivapalan, M., and Gray, W. G.: A unifying framework for watershed thermodynamics: constitutive relationships, Adv. Water Resour., 23, 15–39, 1999.

Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M.: Conservation equations governing hillslope responses: Exploring the physical basis of water balance, 36(7), 1845–1863, doi: 10.1029/2000WR900066, 2000.

Reggiani, P. and Schellekens, J.: Modelling of hydrological responses: the representative elementary watershed approach as an alternative blueprint for watershed modelling, Hydrol. Process., 17, 3785–3789, 2003.

Reggiani, P. and Rientjes, T. H. M.: Flux parameterization in the representative elementary watershed approach: Application to a natural basin, Water Resour. Res., 41, W04013, doi: 10.1029/2004WR003693, 2005.

Rinaldo, A., Rigon, R., Banavar, J. R., Maritan, A., and Rodriguez-Iturbe, I.: Evolution and selection of river networks: Statics, dynamics, and complexity, Proceedings of the National Academy of Sciences, 111(7), 2417–2424, doi: 10.1073/pnas.1322700111, 2014.

Rodriguez-Iturbe, I., Rinaldo, A.: Fractal River Basins: Chance and Self- Organization. Cambridge University Press, Cambridge, UK, 2001.

Roux, H., Labat, D., Garambois, P.-A., Maubourguet, M.-M., Chorda, J., and Dartus, D.: A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments, Nat. Hazards Earth Syst. Sci., 11, 2567-2582, doi:10.5194/nhess-11-2567-2011, 2011.

Safari, A., De Smedt, F., and Moreda F.: WetSpa model application in the Distributed Model Intercomparison Project (DMIP2), Journal of Hydrology, 418, 78–89, doi:10.1016/j.jhydrol.2009.04.001, 2012.

Salvadore, E.: Development of a flexible process-based spatially-distributed hydrological model for urban catchments, PhD thesis, Vrije Universiteit Brussel, Belgium, 2015.

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.

Savenije, H. H. G.: HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)", Hydrol. Earth Syst. Sci., 14, 2681-2692, doi: 10.5194/hess-14-2681-2010, 2010.

Schaefli, B., Hingray, B., Niggli, M., and Musy, A.: A conceptual glacio-hydrological model for high mountainous catchments, Hydrol. Earth Syst. Sci., 9, 95-109, doi: 10.5194/hess-9-95-2005, 2005.

Schaefli, B. and Gupta, H. V.: Do Nash values have value? Hydrological Processes, doi: 10.1002/hyp.6825, 2007.

Schaefli, B. and Zehe, E.: Hydrological model performance and parameter estimation in the wavelet-domain, Hydrol. Earth Syst. Sci., 13, 1921-1936, doi: 10.5194/hess-13-1921-2009, 2009.

Schaefli, B., Harman, C. J., Sivapalan, M., and Schymanski, S. J.: HESS Opinions: Hydrologic predictions in a changing environment: behavioral modeling, Hydrol. Earth Syst. Sci., 15, 635-646, doi: 10.5194/hess-15-635-2011, 2011.

Schaefli, B. and Huss, M.: Integrating point glacier mass balance observations into hydrologic model identification, Hydrol. Earth Syst. Sci., 15, 1227-1241, doi: 10.5194/hess-15-1227-2011, 2011.

Schymanski, S. J.: Optimality as a concept to understand and model vegetation at different scales, Geography Compass, 2(5), 1580–1598, doi: 10.1111/j.1749-8198.2008.00137.x, 2008.

Schymanski S.J., Sivapalan M., Roderick M.L., Hutley L.B., Beringer J.: An optimality-based model of the dynamic feedbacks between natural vegetation and the water balance. Water Resources Research 45:W01412, doi: 10.1029/2008WR006841, 2009.

Semenova, O., Lebedeva, L., and Vinogradov, Y.: Simulation of subsurface heat and water dynamics, and runoff generation in mountainous permafrost conditions, in the Upper Kolyma River basin, Russia. Hydrogeology Journal Official Journal of the International Association of Hydrogeologists, doi: 10.1007/s10040-012-0936-1, 2013.

Semenova, O. M., Lebedeva, L. S., Nesterova, N. V, and Vinogradova, T. A.: Evaluation of short-term changes of hydrological response in mountainous basins of the Vitim Plateau (Russia) after forest fires based on data analysis and hydrological modelling, 3711575194(10), 157–162, doi: 10.5194/piahs-371-157-2015, 2015.

Semenova O. and Beven, K. J.: Barriers to progress in distributed hydrological modelling, Hydrol. Process. doi: 10.1002/hyp.10434, 2015.

Sivapalan, M.: Pattern, process and function: elements of a unified theory of hydrology at the catchment scale, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., vol. 1, Wiley, Chichester, 193–220, 2005.

Smith, M. B., Koren, V., Zhang, Z., Zhang, Y., Reed, S. M., Cui, Z., Moreda, F., Cosgrove, B. A., Mizukami, N., Anderson, E. A., and DMIP 2 Participants: Results of the DMIP 2 Oklahoma experiments, Journal of Hydrology, doi: 10.1016/j.jhydrol.2011.08.056, 2012.

Tian, F., Hu, H., Lei, Z., and Sivapalan, M.: Extension of the Representative Elementary Watershed approach for cold regions via explicit treatment of energy related processes. Hydrol. Earth Syst. Sci., 10(5), 619–644, doi: 10.5194/hess-10-619-2006, 2006.

Vidal, J. P., Moisan, S., Faure, J. B., and Dartus, D.: River model calibration, from guidelines to operational support tools, Environmental Modelling and Software, doi: 10.1016/j.envsoft.2006.12.003, 2007.

Viglione, A., Merz, R., Salinas, J. L., and Bloschl, G.: Flood frequency hydrology: 3. A Bayesian analysis. Water Resources Research, doi: 10.1029/2011WR010782, 2013.

Vinogradov, Y. B.: The "Hydrograph GGI-90" model and its application for mountain basins, Hydrology in Mountainous Regions. I- Hydrological Measurements; the Water Cycle (Proceedings of two Lausanne Symposia, August 1990), IAHS Publ. no. 193, 1990.

Vinogradov, Y. B., Semenova, O. M., and Vinogradova, T. A.: An approach to the scaling problem in hydrological modelling: The deterministic modelling hydrological system. Hydrological Processes, doi: 10.1002/hyp.7901, 2011.

Viviroli, D., Mittelbach, H., Gurtz, J., and Weingartner, R.: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results, J. Hydrol., 377, 208–225, doi:10.1016/j.jhydrol.2009.08.022, 2009.

Vogel, H. J. and Ippisch, O.: Estimation of a critical spatial discretization limit for solving richards' equation at large scales, Vadose Zone Journal, 7, 112-114, 10.2136/vzj2006.0182, 2008.

Weiler, M. and Beven, K.: Do we need a Community Hydrological Model? Water Resources Research, doi: 10.1002/2014WR016731, 2015.

Westhoff, M. C. and Zehe, E.: Maximum entropy production: can it be used to constrain conceptual hydrological models?, Hydrol. Earth Syst. Sci., 17, 3141–3157, doi: 10.5194/hess-17-3141- 2013, 2013.

Wienhofer, J. and Zehe, E.: Predicting subsurface stormflow response of a forested hillslope - the role of connected flow paths, Hydrol. Earth Syst. Sci., 18, 121-138, 10.5194/hess-18-121-2014, 2014.

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.

Zehe, E. and Bloschl, G.: Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, Water Resources Research, doi: 10.1029/2003WR002869, 2004.

Zehe, E. and Sivapalan, M.: Towards a new generation of hydro- logical process models for meso-scale: an introduction, Hydrol. Earth Syst. Sci., Special Issue: Towards a new generation of hy- drological process models for meso-scale, 10, 1–7, 2007.

Zehe, E., Blume, T., and Bloschl, G.: The principle of "maximum energy dissipation": a novel thermodynamic perspective on rapid water flow in connected soil structures, Philos. T. R. Soc. B, 365, 1377–1386, doi:10.1098/rstb.2009.0308, 2010.

Zehe, E., Ehret, U., Blume, T., Kleidon, A., Scherer, U., and Westhoff, M.: A thermodynamic approach to link self-organization, preferential flow and rainfall–runoff behaviour, Hydrol. Earth Syst. Sci., 17, 4297–4322, doi: 10.5194/hess-17-4297-2013, 2013.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, Hydrol. Earth Syst. Sci., 18, 4635-4655, doi: 10.5194/hess-18-4635-2014, 2014.

Zeinivand, H. and De Smedt, F.: Hydrological modeling of snow accumulation and melting on river basin scale. Water resources management, 23(11), pp.2271-2287, 2009.

Zeinivand, H. and De Smedt, F.: Prediction of snowmelt floods with a distributed hydrological model using a physical snow mass and energy balance approach. Natural hazards, 54(2), 451-468, 2010.

Dear Editor Prof. Zehe,

I thank you very much for giving me enough time to rework my paper. I have prepared the revised version of the opinion paper. This revised version is also refined and enhanced by Hoshin Gupta. I had comments of 3 referees, your comments as the editor and the comments of 4 other researchers left on the HESS website which I accepted all of them and used them to improve my work. I must say I could not do this work without the comments and encouraging emails which I have received during one year being involved with this paper. The paper received comments and positive remarks of 25 hydrologists, perhaps due to its clear message. To some scientists like Prof. Hoshin Gupta and Prof. Florimond De Smedt and the three referees (Prof. Beven, Prof. Montanari, and Prof. Schaefli) and you the editor Prof. Zehe, I owe a lot. Their comments were highly significant for the improvement of the work.

In my opinion, the main and major comments, which I addressed them in the paper and used those to improve my work, were these:

1. As it was commented by Montanari and Schaefli, the paper was pessimistic on auto optimization I moderated my statements and also I wrote about the advantages of auto calibration. More than 15 lines are discussing the auto calibration now (lines 356-374).
2. The paper had few examples of physical models, I improved this very much by adding many examples of physics based models. Some of the examples present no calibration in physical based distributed models, some mention limited calibration or just parameter adjustments, and some are the examples of expert knowledge in calibration or parameter specification. For this issue, in addition to the previous citations, I cited and discussed 29 papers as references. All reviewers and the editor had asked me to mention some examples of physics based models. So I did my best to fill the gap. Lines from 84 to 120, then from 149 to 165 are new.
3. I wrote a full new text (whatever I could) about parameter allocation. I owe this to the referee Prof. Schaefli who mentioned several good questions. So while I tried to answer those questions I found out that I have extended my work several pages more! I am happy that I could improve the paper in this regard (more than 135 lines are added for parameter allocation). It was much longer, but fortunately I could decide to delete 3 long paragraphs upon Hoshin Gupta's suggestion.
4. I had several long email conversations with Prof. Beven which I learned a lot through those emails and his thoughtful comments. In most of those emails, he asked me "how it works?". I really did my best to write my paper in this direction to have an answer for his question. I do not know if I was successful, but I have to say the entire Section 5 (196 lines) might provide an answer for this question. Trying to answer this question, I improved and extended the paper very much, it became twice as before. So, I really owe Keith Beven for making the review procedure so challenging for me.
5. I had the feeling that a modeling based upon a thermodynamic approach is the right track which I should emphasize it but I was not sure until receiving the editor's comment. So an important change in my revised version is the emphasis on energy centered hyrological modeling. Editor comments really helped me a lot to make a much better paper.
6. The first version had nothing about data and measurement. Prof. Beven and Dr. Sheikh pointed out this gap, so, I wrote a paragraph to feel this gap (lines 506 to 511, also please see lines 77 to 81)
7. Apart from the comments, some newer approaches like REW modeling, Behavioral modeling, optimality approach, models of everywhere, and community model were discussed (they are discussed in different parts of the paper but mainly in section 5, in particular subsection 5.3, e.g., lines 464-486). I wrote my opinion about the future of hydrological modeling in an original example which I have explained it as spherical jigsaw puzzle modeling (subsection 5.3).

8. I also wrote more about the wrong physics being used in our modeling (327-346 from the first version, and 512-516 of the revised version).

I really appreciate the very good choice of appointing the right referees for this work. I have to say the referees and the editor comments made the work very much better. The mentioned gaps were filled in, as so the length of the paper increased more than twice. While the previous submission was 428 lines, the new version is 914 lines (despite being shortened by Hoshin). The new version has 114 references, while the first submission had only 40 references. I made a marked-up manuscript too. More detail is written as the marked-up comments.

The changes according to each reviewer, separately:

1. Prof. Beven: he asked me a revised version after a long email discussion. I tried to use all his comments in different parts of the paper. But mainly these lines are directly related to Beven's comments: 77-84, 347-577. In the marked up file, I have commented in different parts, for example, I deleted the GLUE example which was correctly mentioned as a bad practice. I gave a special attention to the model of everywhere and learning process in the jigsaw puzzle example, as well as several other significant opinions of Prof. Beven briefly mentioned (e.g. equifinality, GLUE, modeling protocol, self-organized dissipation of singular events, hyper resolution and community model, closure problem, wrong physics, uniqueness of place, etc.).
2. Prof. Montanari: he recommended me to consider 3 corrections in my paper, he clearly told me how to do them (It is appreciated). Lines 64-65 (trial and error for initial values), lines 84-120 (knowledge based optimization and physics based modeling examples), line 356-374 (advantages of auto calibration) .Prof. Montanari also asked me to clarify my idea about calibration, which I did this very clear now. I can say one third of the paper now proves how I think of calibration but please see lines of 356-374, several other sentences talking about limited calibration, parameter adjustments, and calibration not only according to local data but also in conformity with the higher level water balances as well as organizing principles, etc. I also wrote the calibration is unavoidable (line 357).
3. Prof. Schaefli: she posed several clarifying questions which I tried to address them all. The entire subsections 5.1 and 5.2 are written in response to her comments. By the way, I built a close discussion between my opinion and her opinion presented in Schaefli et al. 2011. Schaefli had also emphasized on comments of Montanari.
4. Prof. Zehe: I added many examples of physics based modeling to over shadow some examples of conceptual bucket models. So, almost 80% of the examples are now of physics based models. These are some of the models: hydrograph model, TOPKAPI, CATFLOW, MIKE SHE, WetSpa, WetSpa-Python, MARINE, THREW, etc. I had a special emphasize on new works which consider energy balances too. This can be seen in the entire marked-up file. Although, while discussing my opinions often I mentioned other opinions too, but because, I did not see my message something against the common practice in hydrology so the paper did not become much in dialectic sense, but I am convinced it has clear messages without disregarding other opinions.
5. Prof. Sadeghi and Dr. Sheikh: I avoided to use the word "conceptual" in the abstract, the "empirical" (proposed by Hoshin Gupta) serves better. I wrote a paragraph about data and measurements (506-511).

At, the end again I thank you very much for all your guidance and support, and I hope this version suits the high level journal HESS. I also appreciate the referee’s valuable comments. I am ready to improve the manuscript more as much as it needs.

Best regards,

Abdolreza Bahremand