Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment–MOPEX. IAHS Publ. 307, 2006.

Has basin-scale modelling advanced beyond empiricism?

C. MICHEL[1], C. PERRIN[1], V. ANDRÉASSIAN[1], L. OUDIN[2] & T. MATHEVET[1]

1 *Cemagref, Hydrosystems and Bioprocesses Research Unit, Parc de Tourvoie, BP 44, 92163 Antony cedex, France*
claude.michel@cemagref.fr

2 *Université Pierre-et-Marie Curie, UMR SISYPHE, 4 place Jussieu, 75252 Paris cedex 05, France*

Abstract Today, due to the increasing availability of spatial representations of our environment, there seems to be a growing feeling that hydrological modelling can eventually produce efficient distributed physically-based rainfall–runoff (RR) models. However, the improvement brought by such models is still to be demonstrated. Several authors have already sounded the alarm bell on the development and application of these models, apparently to no avail. In this paper, we argue that hydrological modelling has not gone far beyond an empirical view of the way basins transform rainfall into streamflow at their outlet. We address simple questions relative to: (a) the unit physical object that should be represented by RR models; (b) the actual limits of lumped RR models; (c) the impact of the time step on the structure of RR models; (d) the necessity of *a priori* conceptualization in the design of RR model structures; and last (e) what defines a "good" model. We are convinced that, up to now, the empirical modelling approach has not been studied with the rigour that would help to discriminate between the numerous mathematical tools at the hydrologist's disposal to build model structures. Thus, the purpose of this paper is to recall that before considering complex models, a lot remains to be done in the area of simple lumped models.

Key words catchments; empiricism; lumped modelling; rainfall–runoff modelling; time-step

INTRODUCTION

In this paper the only study variable that will be considered is the water flow observable through a given section of a stream. The prediction of streamflow requires modelling the transformation of precipitation that has fallen over the basin area upstream of the point of interest. Two opposed approaches are available to build such a model: the upward approach (also called bottom-up, reductionist or mechanistic) and the downward approach (also called top-down, empirical or non-reductionist). This distinction, which reflects a well-known debate of scientific methodology (see for example von Bertalanffy, 1968), was discussed at length by Klemeš (1983), who defines the downward approach as the route that "*starts with trying to find a distinct conceptual node directly at the level of interest (or higher) and then looks for the steps that could have led to it from a lower level*". Conversely, in the upward approach that dominates modern science, basin properties are considered to be a summation of the hillslope and stream channel properties, at all scales.

To assist hydrologists in their model-building enterprise, today there is an unprecedented amount of data and information on the catchment, supplied by various

measuring devices ranging from ground networks to satellites, at various time and space scales, arranged into powerful databases and displayed by user-friendly geographical information systems (see e.g. Babovic, 2005). Although most of the hydrologically-relevant information remains unknown to us (especially subsurface basin characteristics), this wealth of information appeals like a mirage to those of the reductionist modellers who consider that every cube of earth can be physically described and modelled, at any resolution in time and space. This may be possible in the idealized conditions of laboratories where parameter values and boundaries conditions can be measured and recorded to allow the application of the well-established concepts of hydrodynamics (see Beven, 2001, for examples).

However, such conditions do not exist in the real world and these concepts are inapplicable to solving the problem of rainfall–runoff transformation at the catchment scale. Indeed, in spite of the apparent plethora of existing data, these requirements cannot be satisfied over natural catchments because of the impossibility of monitoring the huge and intricate three-dimensional (3-D) landscape involved in the rainfall–flow transformation. To cope with our ignorance of the key basin characteristics, reductionist modellers are forced to introduce hypotheses and subjective parameters that can be far from reality: their physically-based models then fall back to the level of simple conceptual or empirical models.

Despite the alarm bells sounded by authors such as Beven (1989), Bergström (1991), Jakeman & Hornberger (1993) and Young (1998), many hydrologists believe that reductionist modelling is the best way, i.e. the way that will yield significant progress in rainfall–runoff (RR) modelling (see e.g. Loague & VanderKwaak, 2004). Our view is that, even though some progress has been made over the past decades in the understanding of the hillslope processes, little advance has been made in catchment scale RR modelling (see e.g. Sivapalan, 2003). And, we think that the best way to advance our hydrological modelling, is through an *empirical lumped approach* (following a downward approach) very similar to the data-based modelling approach advocated by Young (2003).

In this paper, we discuss several of the generally accepted ideas in hydrological modelling and some of the fundamental questions that today seem insufficiently or inadequately dealt with by the hydrological community. Our paper is organized around five successive questions: we first discuss the unit physical object on which RR models should be based. Second, we examine the actual limitations of lumped models, in comparison to the distributed models. Third, we discuss issues of the time step in modelling. Fourth, we contest the widely accepted belief that conceptualization is a prerequisite to model building, and advocate data-based approaches. Fifth, we expose the desirable statistics on which we can base our judgement to consider a model as "good".

WHAT UNIT PHYSICAL OBJECT SHOULD RAINFALL–RUNOFF MODELS BE BASED ON?

Unfalsifiable distributed models

The object of the study of hydrology is the catchment: an extraordinarily complex, 3-D system (a catchment is even 4-D as it is not always stationary), subjected to a similarly

complex climatic forcing, and producing visible (streamflow) and invisible (atmospheric and underground water fluxes) outputs. The constituent unit of a hydrological model depends on the modelling approach: the elementary unit will be a grid cell in a distributed model, a basin in a lumped model.

The problem we see with the use of grid cells as elementary units, is that the modelling of these components is necessarily hypothetical. Making each constituent unit work properly is a necessary *but not sufficient* condition to make the whole model mimic the actual basin functioning. The construction of the full model is not a simple matter of straightforward aggregation of all the constituent units (see e.g. Sivapalan, 2003). The laws that govern the interactions between these elementary units remain to be discovered. Indeed it is necessary to figure out how fluxes originating from one element are integrated by the adjoining elements and how the state variables of one block interact with those of the neighbouring blocks. Thus, there are numerous mingled hypotheses, those made about the constituent blocks and those made about the interactions between them. Only when the whole structure has been devised does it become possible to confront the single model output to reality, i.e. the observed discharge at the basin outlet.

The problem with physically-based models is that all constituent hypotheses *taken separately* are reasonable. Thus, there cannot be any strong refutation of the whole construction, and if a few tuning parameters are left to the model user, it will be easy to adapt the whole construction to achieve a good simulation of flows. A scientific approach would require that one questions the model, but the numerous and varied hypotheses cannot be challenged individually. We have thus obtained a model that yields reasonable results, but that is not questionable. Clearly, such a model is not falsifiable (i.e. it cannot be corroborated or refuted in the sense of hypothesis testing) whereas it should be (see e.g. Oreskes *et al.*, 1994; Refsgaard & Henriksen, 2004). It is all the more unfalsifiable that it is generally applied on a single basin with no consideration for possible falsification on other basins.

Model complexity

In addition to these problems, the aggregation of grid cells systematically raises the issue of model complexity. As noted by Sivapalan (2003), the complexity of the resulting model can be higher than the complexity of individual units, and this will result in problems of over parameterization. However, in the bottom-up approach, today there is no guideline to rationally reduce this complexity during the aggregation process. This is probably partly due to the "frustration" that such simplifications may cause, as noticed 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*".

This upward approach for model construction therefore seems unable to design parsimonious models. However, many results indicate that a limited number of free parameters (less than six) are sufficient to simulate the RR relationship (see e.g. Jakeman & Hornberger 1993; Perrin *et al.*, 2001; Young, 2003). Simple models are often as good as complex models in simulating the RR relationship. As mentioned by

Martin (1996), "*the prediction obtained with a complex model often points to a simpler model which could have been used in the first place*".

Going back to the basin

Our overriding feeling is that there is a need to go back to the basics and to study as a whole the object responsible for the observed flow, i.e. the basin. During the past 20 years, hydrologists have considered that the necessary basis for any modelling effort was to look at the basin as a sum of grid cells (or sub-basins). We believe they have overlooked the actual challenges presented to our science. Before partitioning the basin into smaller areas, it seems advisable to look at the entire object itself and to try and get a general idea of its functioning. To put it another way, the first action to be taken is to improve lumped models. Such models are simple, limit the problems related to undue complexity and can be easily tested and falsified.

CAN WE ACHIEVE FURTHER PROGRESS WITH LUMPED MODELS?

Why do we need lumped models?

Lumped modelling searches for the mechanisms relevant *at the basin scale* in order to describe the *overall hydrological behaviour* of a catchment. As an individual concerned with his health will first consult a physician who will handle him as a person (and not as a sum of cells), we believe that any hydrological study should start with an investigation of the basin as a single functional entity. The empirical lumped approach recognizes the impossibility of monitoring the great complexity of the 3-D geographical domain involved in the rainfall–flow transformation. Thus, it chooses to give a central part to the data in the design of a model structure, which will reveal itself by successive hypothesis testing, avoiding constraints in the model development by employing (often wrong) preconceived ideas.

In a lumped context, it seems obvious that the model cannot be other than an abstract representation of reality. The links between the parameters of this abstract structure with the physical reality cannot be set explicitly and *a priori*: they remain implicit, to be discovered *a posteriori*. The reason is that phenomena observable at the plot scale are no longer relevant when applied at the basin scale in a lumped mode, in the same way that the mechanisms of neuronal functioning are not directly relevant to explain human behaviour at the individual's scale. One has to think of entirely new tools suitable for basin-scale modelling, raising basic questions such as:

(1) Is a subtraction of rainfall input (infiltration model) more relevant than a multiplication by a number less than one (runoff coefficient)?
(2) Is it necessary to split the outflows into more than two components?
(3) Is the bucket concept with an upper bound to its effective inputs an efficient way to model the RR transformation at the basin scale?

Naturally, these simple questions may seem disturbing, as distributed modelling has pre-designed answers to them, but lumped modelling has none. We believe that this is the reason why lumped modelling is all the more suitable for scientific

investigation: it raises questions, leaves room for doubting, and allows easy tests of hypotheses.

Lumped models as scientific investigation tools

To judge the suitability of lumped models, it can be interesting to look at them from an uncertainty analysis point of view, which is indeed an important aspect of hydrological modelling. Up-to-date techniques such as Kalman filtering are currently used to track model states and parameters. Which models lend themselves to these uncertainty analyses? Simple lumped models! For instance, in their demonstration of the use of sequential data assimilation, Moradkhani *et al.* (2005) resort to the HyMOD model, a simple five-parameter model composed of the probability distributed moisture model (Moore, 1985) and of linear reservoir routing. Instead of disappearing from the scientific literature, these simple models are employed more often than before (see e.g. Sivapalan *et al.*, 2003). The reason could be that lumped conceptual models are subject to more effective scientific scrutiny than distributed models whose scientific foundation is taken for granted. However, a lot of work remains to be done to discriminate among the numerous models published in the hydrological literature. Lumped modelling is definitely not a thing of the past and we think it can be the cornerstone on which future hydrology will rise.

DOES THE TIME STEP OF A MODEL MATTER?

Surprisingly, the time step of model inputs and outputs is often considered as incidental information. Very few hydrologists (see e.g. Jothityangkoon *et al.*, 2001; Eder *et al.*, 2003; Mouelhi, 2003) have questioned the influence of the time step on model structure. Modellers generally present their models as if the time step of its functioning was just a matter of data availability, or an exclusively computational question. In fact, the influence of the time step on model structure is far more important than the influence of spatial discretization. If a model works satisfactorily in a lumped mode, it will similarly do so when used in a semi-distributed mode. In contrast, a model successful at a monthly time step may yield poor performances when it is run at a daily time step. Each time scale should have its proper RR modelling characteristics. Generally, the longer the time step, the less complex the model structure and the smaller the number of free parameters. Therefore, it is far simpler to develop an annual model than a daily model. However, the link between model structures at different time steps is not straightforward (Jothityangkoon *et al.*, 2001; Mouelhi, 2003).

IS PROCESS DESCRIPTION THE GOAL?

"Processes" at the basin scale

If hydrologists tend to relinquish the strict physically-based approach for model development, they generally stick to the necessity of having a clear view of the actual

processes taking place in a basin (a predefined "perceptual" model). They fear that if a model does not explicitly take into account the well-known small-scale processes, this model will lose its credibility, and if it still happens to work, it will be so for the "wrong reasons". This is quite a weird idea, and this argumentation does look circular. Generally, modelling hypotheses are justified by the fact that the model based on them achieves better results than another model that rejects them. Processes are very complicated, intertwined and highly variable over the basin (see e.g. Joerin *et al.*, 2005). To the basin-scale modeller, the important thing is the end result of this small-scale complexity, i.e. the discharge measured at the gauging station. The problem to be solved is to find the combination of mathematical equations that is the most successful in predicting the discharges.

Flows and other hydrological variables

Once a model has been developed based on a conceptualization approach, the whole structure is often taken for granted, i.e. every part of it is believed to have its counterpart in the real basin (see e.g. Günter *et al.*, 1999). As far as lumped empirical modelling is concerned, such a consideration has no sound foundation because the only concern of lumped modelling is to produce discharges that are close to the observed discharges. Given lumped models generally have a single target variable (the discharge), there is no point at all in trying to find *a posteriori* links between model states and other variables observed on the basin. As pointed out by Beven (2001), it is misguided to give more consideration to searching for "explanatory depth" (i.e. internal physical realism) in the model than to searching for improved "predictive power" (i.e. the capacity to satisfactorily simulate the target variable).

As an afterthought, some modellers have tried to include secondary modelling objectives, i.e. secondary outputs (e.g. soil moisture, piezometric level, etc.), in order to constrain those parameters that appeared too poorly identified (see e.g. Lamb *et al.*, 1998). We believe that if two or more outputs are to be predicted, the whole model development has to be started again. This implies the definition of new objective functions and addressing difficult issues such as the guidelines for dealing with Pareto optima. It is an altogether new problem that requires a lot more data than are generally available to hydrologists.

The reason for distrusting the internal functioning of a lumped model is that aggregation of processes is unavoidable. In a parsimonious model developed following the Okham's razor rule (i.e. discarding model components not leading to a clear advantage in terms of performance), processes elicited by model development can overlap many elementary physical processes singled out in field studies. A lumped model should not and cannot be a compendium of all the processes that can be observed in the real world. It is rather an abstract structure that is only justified by its ability to simulate satisfactorily streamflow at the outlet of any basin, provided that its parameters are adjusted to suitable values. For these reasons, we consider that the prediction of discharges at the basin outlet is the primary objective of RR models and, unfortunately, there is still a lot of work to efficiently simulate these discharges.

"GOOD" OR "BETTER" MODELS?

Limits in testing a single model

It is often said that when the Nash & Sutcliffe (1970) criterion is above 0.8, the model fit can be considered as good. It is easy to show the limitations of such a statement. That a model yields a Nash-Sutcliffe criterion equal to 0.8 on a given basin, for a given period, is in no way a validation: it is rather a judgement on the basin and on the period. If it happens that this period has been particularly wet, there is a high probability of obtaining poorer results in a drier period. If the same model were to be applied to other basins, even within the same region, there is an even higher probability of obtaining a range of Nash-Sutcliffe criteria spanning from low to high values. Therefore, model performance on a single basin and a single period provides little information on the actual model value.

Also note that, if the Nash-Sutcliffe criterion is widely used as a derivative of the root mean square error (RMSE), it nevertheless does not constitute an absolute reference. The model user may prefer other criteria such as the mean absolute error calculated on discharges or on the logarithms of discharges and so on.

Sometimes, a model is considered as "good" when the RMSE obtained in simulation ("validation") on a period different from the calibration period is not clearly larger than the RMSE obtained in calibration. But actually both RMSEs are generally quite close. In that sense, should all models be considered as satisfactory? Or is it better to acknowledge that there cannot be any absolute measure of the quality of a model and, consequently, there is no possibility to consider a model as "good".

Model comparison as objective assessment

Many years ago, one of the founding fathers of hydrological modelling, Ray Linsley (1982), pointed out that "*almost any model with sufficient free parameters can yield good results when applied to a short sample from a single basin, [so that] effective testing requires that models be tried on many basins of widely differing characteristics, and that each trial cover a period of many years*". We concur with this statement: first, we think that a model can be properly assessed only if it is applied on a large sample of basins representing a wide range of hydro-climatic conditions (see Mathevet *et al.*, 2006). This gives clear ideas on its robustness and generality. Second, we think that a model cannot be considered as *good* but only as *better* than another model. Therefore, the model should be assessed comparatively to another model, as also advocated by Seibert (2001). If $NS_A(k)$ is the Nash-Sutcliffe criterion obtained by model A on basin k and if $NS_B(k)$ is the Nash-Sutcliffe criterion obtained by model B on the same basin, then a statistically founded opinion can be reached if a statistical test can demonstrate that there is a high probability that the mean of $NS_A(k)$, $k = 1, 2, \ldots, N$ be different from the mean of $NS_B(k)$, $k = 1, 2, \ldots, N$ over the whole sample of N basins. Only such a comparative framework on large samples of basins will give reliable conclusions on the actual merits of RR models.

DISCUSSION: WHAT IS THE STATE OF THE ART OF LUMPED RAINFALL–RUNOFF MODELLING?

Since a lumped model is unable to reproduce all processes taking place in a basin, its simulations are bound to be in error (actually, this statement applies to all kinds of models). It is rare to obtain a Nash-Sutcliffe criterion larger than 0.9. This means that the RSME is commonly greater than 0.32 times the standard deviation of observed discharges. Thus, there is plenty of room for progress.

It is unlikely that a unique structure can emerge as the best model, just because there is no physical reason for that. For the time being, there are a lot of competing structures that are approximately equivalent, although some are definitely inferior (e.g. Perrin *et al*., 2001). However, the hydrological community has not yet recognized which models are in the upper class and which ones are to be forsaken. The reason is that, in spite of several efforts to compare model structures (e.g. WMO, 1975; Smith *et al.*, 2004), a lot of questions have yet to be answered. In particular, there is no indication as to which tools (unit hydrographs, linear and nonlinear stores, etc.) are the best suited to the loss and transfer sub-models. In order to represent the hydrological behaviour of a basin, it seems as if the choice between several unit hydrographs, different sorts of routing reservoirs, etc., is a matter of sheer convenience. It is probably not.

CONCLUSION

In this paper, we have tried to answer some simple questions on the present state of RR modelling. We have the overriding impression that lumped empirical modelling has so far not received the attention it deserves and requires. At the basin scale, it seems necessary to grow out of the analysis of small-scale processes that cannot be accounted for in a tractable way, and to explore the emergent properties of the catchment behaviour: this quest must necessarily start with simple lumped models. Although most of the constituent tools needed to represent the emerging properties of a drainage basin are already known, hydrologists have not yet identified which are the most promising ones and understood their roles in the whole model structure. We definitely think that at this stage of hydrological modelling, empirical research remains the most promising approach. Large databases and model comparisons should be key tools to support future progress.

REFERENCES

Babovic, V. (2005) Data mining in hydrology. *Hydrol. Processes* **19**(7), 1511–1515.

Bergström, S. (1991) Principles and confidence in hydrological modelling. *Nordic Hydrology*, **22**, 123–136.

Beven, K. J. (1989) Changing ideas in hydrology—the case of physically-based models. *J. Hydrol.* **105**, 157–172.

Beven, K. (2001) On explanatory depth and predictive power. *Hydrol. Processes* **15**(15), 3069–3072.

Eder, G., Sivapalan, M. & Nachtnebel, H. P. (2003) Modelling water balances in an Alpine catchment through exploitation of emergent properties over changing time scales. *Hydrol. Processes* **17**(11), 2125–2149.

Güntner, A., Uhlenbrook, S., Seibert, J. & Leibundgut, C. (1999) Multi-criterial validation of TOPMODEL in a mountainous catchment. *Hydrol. Processes* **13**(11), 1603–1620.

Jakeman, A. J. & Hornberger, G. M. (1993) How much complexity is warranted in a rainfall–runoff model? *Water Resour. Res.* **29**(8), 2637–2649.

Joerin, C., Beven, K. J., Musy, A. & Talamba, D. (2005) Study of hydrological processes by the combination of environmental tracing and hill slope measurements: application on the Haute-Mentue catchment. *Hydrol. Processes* **19**(16), 3127–3145, doi:10.1002/hyp.5836.

Jothityangkoon, C., Sivapalan, M. & Farmer, D. L. (2001) Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. *J. Hydrol.* **254**(1–4), 174–198.

Klemeš, V. (1983) Conceptualization and scale in Hydrology. *J. Hydrol.* **65**, 1–23.

Lamb, R., Beven, K. and Myrabø, S. (1998) Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. *Adv. Water Resour.* **22**(4), 305–317.

Linsley, R. K. (1982) Rainfall–runoff models—an overview. In: *Proc. Int. Symp. Rainfall–Runoff Modelling* (ed. by V. P. Singh), 3–22. Water Resources Publications, Littleton, Colorado, USA.

Loague, K. & VanderKwaak, J. E. (2004) Physics-based hydrologic response simulation: platinium bridge, 1958 Edsel, or useful tool. *Hydrol. Processes* **18**(15), 2949–2956.

Martin, P. H. (1996) Physics of stamp-collecting? Thoughts on ecosystem model design. *Sci. Total Environ.* **183**, 7–15.

Moore, R. J. (1985) The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.* **30**(2), 273–297.

Moradkhani, H., Hsu, K-L., Gupta, H. & Sorooshian, S. (2005) Uncertainty assessment of hydrological model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.* **41**, W05012, doi:10.1029/2004WR003604.

Mouelhi, S. (2003) Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. PhD Thesis, ENGREF, Cemagref Antony, France (in French).

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models. Part I—A discussion of principles. *J. Hydrol.* **27**(3), 282–290.

Oreskes, N., Shrader-Frechette, K. & Belitz, K. (1994) Verification, validation and confirmation of numerical models in the Earth Sciences. *Science* **263**, 644–646.

Perrin, C., Michel, C. & Andréassian, V. (2001) Does a large number of parameters enhance model performances? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.* **242**, 275–301.

Refsgaard, J. C. & Henriksen, H. J. (2004) Modelling guidelines—terminology and guiding principles. *Adv. Water Resour.* **27**(1), 71–82.

Seibert, J. (2001) On the need for benchmarks in hydrological modelling. *Hydrol. Processes* **15**(6), 1063–1064.

Sivapalan, M. (2003) Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection. *Hydrol. Processes* **17**(3), 1037–1041.

Sivapalan, M., Zhang, L., Vertessy, R. & Blöschl, G. (eds) (2003) Downward approach to hydrological prediction. *Hydrol. Processes* **17**(11), 2099–2326.

Smith, M. B., Seo, D. J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F. & Cong, S. (2004) The distributed model intercomparison project (DMIP): motivation and experiment design. *J. Hydrol.* **298**(1–4), 4–26.

Young, P. C. (1998) Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environ. Modelling Software* **13**, 105–102.

Young, P. (2003) Top-down and data-based mechanistic modelling of rainfall–flow dynamics at the catchment scale. *Hydrol. Processes* **17**(11), 2195–2217.

von Bertalanffy, L. (1968) *General Systems Theory*. Georges Braziller, New York, USA.