13 New Directions for Top-Down Modelling: Introducing the PUB Top-Down Modelling Working Group

DAVID A. POST, IAN G. LITTLEWOOD & BARRY F. CROKE

BACKGROUND

There are two distinct approaches that can be taken when developing a new hydrological (or perhaps any) mathematical model. One can either take a “top-down” approach, or a “bottom-up” approach. Before describing the approach which will be taken by the new *PUB Top-Down Modelling Working Group* (TDWG), a definition is required for top-down modelling, and the ways in which it differs from bottom-up modelling.

Most importantly, the top-down approach is data-based. That is, it involves learning about catchment functioning and deriving a model structure from the available data. In addition, model refinement is carried out through incorporating new processes only where the observed data supports such refinements in the model structure.

Conversely, the bottom-up approach focuses on developing complex models describing small-scale processes operating in a catchment without consideration of the hydrological processes operating at the catchment scale. These small-scale processes are then up-scaled in an attempt to reproduce the hydrological response of the catchment as a whole. Unfortunately, processes which may be dominant at one scale may be much less important or even irrelevant at another scale. Indeed, the hypothesis that *at the whole-catchment scale, much of the complexity needed to model hydrological response at finer scales is unnecessary* is a fundamental tenet of top-down modelling.

Much has recently been written about the top-down approach in the literature, and the interested reader is referred to the special issue of *Hydrological Processes* vol. 17(11) (2003), for more detail. In addition, papers by Littlewood *et al.* (2003), Sivapalan *et al.* (2003a,b), and Sivapalan & Young (2005) provide ample descriptions of the top-down approach and the ways in which it differs from the bottom-up approach. The purpose of this short chapter is to introduce the Top-Down Modelling Working Group (TDWG) which has been formed under the IAHS Predictions in Ungauged Basins (PUB) umbrella and to describe some of the approaches that will hopefully lead to progress in achieving some of the fundamental aims of PUB.

THE TDWG APPROACH

Within the working group it is recognized that different models may be required to address different problems. The extreme top-down approach is counter but highly complementary to the extreme deterministic or bottom-up approach whereby explanation, understanding and predictive capability are sought by including representation of all processes considered relevant. The idea behind bottom-up modelling is that a single deterministic model should be able to address all questions. The TDWG will deal with hybrid approaches, incorporating features of both approaches but with an emphasis on the top-down end of the spectrum of hybrid models. In some ways, the TDWG aims to take the best ideas from bottom-up modelling and incorporate them into the top-down approach. For example, following the top-down approach, one can easily develop a model which provides an adequate fit to the observed hydrological data, but which has little or no physical meaning. Such a model would be of limited use in regionalization studies. The lesson learnt from the bottom-up approach is that the model must represent (or at least be strongly related to) the physical processes occurring in a catchment.

A core tenet of PUB (and therefore the TDWG) is a focus on the reduction of predictive uncertainty. However, before one can reduce predictive uncertainty, one has to first identify the sources of uncertainty. A brief search of the literature will find many examples where not only are the sources of uncertainty not recognized, it seems that the modellers are blissfully unaware that their model contains any uncertainty—certainly no reference is made to it in the paper. Within the TDWG, it is recognized that a range of approaches may be useful, *as long as the modeller is prepared to consider and tackle the sources of uncertainty in their model predictions*.

AIMS OF THE TDWG

The Top-Down Modelling Working Group (TDWG) will examine the use of top-down modelling as a tool for prediction in ungauged basins. This includes surface water and groundwater hydrology, as well as water quality issues (e.g. sediment, nutrient and contaminant, source, transport and deposition). The variables of interest include but are not restricted to the following: precipitation; streamflow; groundwater levels and volumes; hydro-ecological indicators (e.g. biodiversity variables, physical habitats); fluvial mass loads (fluxes) of sediment, material adsorbed on sediment, and solutes (including diffuse pollution components); soil moisture; atmospheric deposition; evaporation; and transpiration.

The overall aim of the TDWG is to produce models which are capable of prediction in ungauged basins. To do this, it will need to:

1. Quantify, and then reduce the level of predictive uncertainty in the current generation of hydrological models;
2. Develop a new generation of hydrological models which are truly “physically-based”, yet parametrically efficient, and thus suitable for application to catchments where little or no hydrological data are available.

Achieving these two rather ambitious aims will require research to be carried out in three focus areas: data analysis, model analysis and model development. Members of the working group are invited to contribute to any or all of these. While initially the data and model analysis areas will be the key foci, as the working group progresses there will be an increasing emphasis on model development.

A. Data analysis

Members of the TDWG working in this focus area will explore available data sets to develop new, innovative process understanding and formulation. This could develop on existing understanding, but at least part of this focus area should start at a 0-th order hypothesis (i.e. relying only on information contained in data sets and not on preconceived notions of hydrological response), and develop process understanding from data (including metadata and soft data) alone. Data sets to be used should cover the entire spectrum including climate, vegetation, topography, soils, geography, groundwater, water quality, biogeochemistry, etc. Local or traditional knowledge of the catchment can be used, but this should be field tested or validated against other available data.

Additionally, research in this focus area will attempt to determine the set of minimum data requirements to determine the key processes operating in a catchment. Further, there will be interaction between this focus area and the model development focus area through comparison of models with available data sets.

The overall aim of this focus area is to determine what new process understanding might be obtained from currently available data sets. This may be obtained through, for example:

1. Catchment inter-comparisons Even highly studied systems can yield new process understanding when approached in a novel way. For example, Post & Jones (2001) carried out a re-analysis of data from some of the most studied sites in the USA: the Hubbard Brook, Coweeta, H J Andrews and Luquillo Long Term Ecological Research (LTER) sites. Through comparing the hydrological response of 18 control basins across these four sites, the authors were able to come to a better understanding of how precipitation, snowmelt and transpiration impacted on hydrological response in a variety of hydro-climatological regimes.

2. Making use of new technologies for collecting data Previously studied catchments may yield new information as a result of the application of new technologies. One of the more obvious examples is remote sensing, where new data on land cover, land use, soil texture, soil moisture, etc., can be obtained for previously studied catchments. As well as terrestrial information, remote sensing can be used to determine river height or even turbidity. For example, Alsdorf & Lettenmaier (2003) use satellite remote sensing to determine the height of the Amazon River. While current resolution of satellite data restrict this method to relatively large rivers, there is every reason to expect that the technique will become applicable to smaller rivers in the future. Similarly, Choubey (1997) used remotely sensed data to determine the turbidity of an inland reservoir. These new techniques allow researchers to obtain data on discharge and water quality that were previously unobtainable. In addition, field deployable auto-analysers are now capable of producing water chemistry data on an hourly or sub-hourly time step. Kirchner (2004) describes the ways in which these new data sources are likely to alter our understanding of catchment processes. These data may be used to calibrate or refine both hydrological and water quality models.

3. Novel techniques for analysing and interpreting data As well as making use of new techniques for collecting data, there are also many opportunities for making use of new techniques for the analysis and interpretation of pre-existing data. Examples include the fractal scaling observed by Kirchner *et al.* (2001), as well as work examining the use of spectral analysis as a tool for interpreting stream tracer

concentrations (Feng *et al.*, 2004). Techniques which are common in other disciplines, but have received little attention in hydrological circles may also be valuable. One such example is data mining, incorporating techniques such as clustering, classification, association rule extraction, and dominant mode analysis (Spate *et al.*, 2003). Making use of new techniques such as these increases our process-based understanding of catchment behaviour. They will also assist with assessing errors in input/output data, and the effect these errors have on model structure and model uncertainty.

4. Determine the set of measurements required to be able to predict the hydrological response of an ungauged catchment While the aim of the TDWG is develop hydrological models which are suitable for application to catchments where little or no hydrological data are available, it is recognized that data of some sort will be required in order to make predictions of hydrological response. These data requirements will differ depending on what aspect of hydrological response is being considered. For example, to predict the daily streamflow hydrograph, we will undoubtedly need daily precipitation measurements, while to predict suspended sediment concentrations we may need land cover data.

Until now, the set of landscape attributes used in regionalization studies has largely been opportunistic, with the attributes chosen simply because the data happened to be available (e.g. Post & Jakeman, 1999). One of the outcomes of the TDWG will be to develop a suite of measurements which are required in order to be able to predict a particular aspect of the hydrological response of an ungauged catchment. This may also include some information about the required accuracy of those parameters. For example, what density of raingauges is required to adequately capture precipitation inputs? What resolution digital elevation model is required? Do we need data on soil depth? soil moisture? etc.

B. Model analysis

Members of the TDWG working in this focus area will study existing models in order to improve these models such that they contribute to our process understanding of hydrological response. Model structure will be assessed through, for example, sensitivity and uncertainty analysis. There will also be a need to test new models developed within the working group (as well as by others) to provide feedback for model development.

The aim of this focus area is to assess the predictive uncertainty in the current generation of hydrological models. This predictive uncertainty arises through, for example:

1. Over-parameterization This may occur due to attempts by the modeller to represent processes which are inappropriate or unimportant at the scale of interest. Hopefully, models which have been developed using the top-down approach will not have this problem, as a fundamental tenet of this approach is to only include parameters where justified by the observed data. However, the temptation is always there to include a process if the modeller believes it is important, regardless of whether or not this is borne out through observation. Model over-parameterization leads to parameter insensitivity, such that there exist multiple local maxima, where any value of

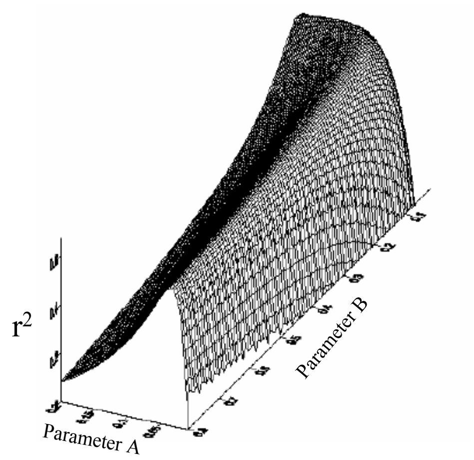

Fig. 13.1 *Example of a poorly defined parameter (B) where any value provides a similar model performance (diagram reproduced with permission from Hreiche et al., 2002; ©2002 iEMSs).*

the parameter leads to an equally good fit. An example of model over-parameterization is provided in Fig. 13.1 (Hreiche *et al.*, 2002).

2. Lack of physical meaning in model parameters In some ways, this is almost the opposite problem to that due to over-parameterization. It is more likely to be found in lumped conceptual models developed using a top-down approach, where model parameters may provide a good fit to the data, but do so for reasons that have nothing to do with the physical processes operating. This will limit the ability of the model to be regionalized. An example is where a conceptual bucket needs to be filled before it overflows and runoff occurs (e.g. the SFB model of Boughton, 1984). From what we know about the movement of water in a catchment, this representation is not accurate, even though it may provide a good fit to the observed hydrograph. A similar problem may also occur in physics-based models, where a physical process may be represented in a way that is not physically meaningful. An example could be using Richards' equation to calculate infiltration in each cell of a distributed model, whereas in reality this relatively slow rate of infiltration is swamped by "hydrological windows" surrounding vegetation where infiltration occurs as macropore flow.

3. Inadequate input data The aim here is to assess the impacts of inadequate input data on the uncertainty associated with model parameters. This focus area is distinct from the data analysis focus area in that here the emphasis is in the impact of inadequate input data on model response. Evaluation of these impacts may then be used to refine the ways in which field data are collected. Errors may arise in both the input data (precipitation data, resolution of the digital elevation model for example), and the output data (streamflow errors for example). These errors may be due to poor spatial resolution (precipitation data again being a good example) or poor temporal resolution (sampling of water quality parameters on a daily or monthly time step for example). Finally, there may simply be a lack of data at the appropriate resolution (measurements of soil moisture, overland flow, or infiltration for example).

C. Model development

The aim of this focus area is to develop new models based on output from the data and model analysis focus areas. Through an understanding of the strengths and weaknesses of existing models, and new process understanding derived from the data, the aim is to develop a new class of models better suited to prediction of hydrological response in ungauged basins. Initially, this will be done through incremental improvements in the current generation of models. However, the longer term aim is to derive a new suite of hydrological models that are physically-based, yet parametrically efficient, and applicable to catchments where we have little or no hydrological data.

Incremental improvements which might occur in the current generation of hydrological models may include for example:

1. Improving the representation of physical processes This may occur through the removal of parameters which are found to be poorly identified through having large uncertainties associated with them. An example of a redundant model parameter is shown in Fig. 13.1. Conversely, Wagener & Wheater (2002) provide an example of a model parameter which has very good identifiability (Fig. 13.2). This parameter has no local maxima, and there is a clearly identified value which provides a better model fit than all other values. Ideally, every model parameter should have a physical meaning and should be able to be measured in the field. In reality, this will not be the case and some form of model parameter calibration will probably need to occur, at least for the current generation of hydrological models.

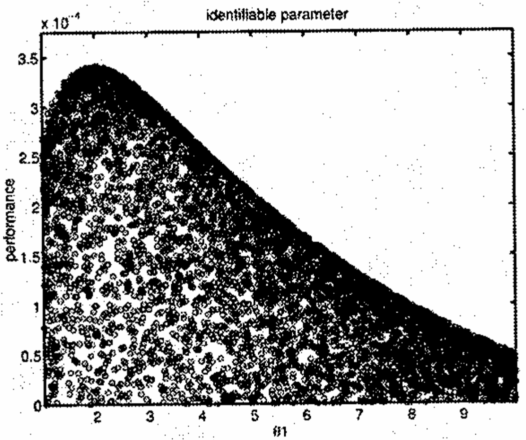

Fig. 13.2 *Example of a well defined parameter. The range of parameter values is shown on the x-axis, with model performance on the y-axis (diagram reproduced with permission, from Wagener & Wheater, 2002; ©2002 iEMSs).*

2. Representing a number of physical processes through one simplified mathematical representation While it is desirable to have a physical meaning associated with each model parameter, it is still likely that this parameter will be representing the effect of a number of physical processes, perhaps occurring simultaneously. In fact, trying to represent every physical process of importance occurring in a catchment will almost certainly lead to model over-parameterization and is therefore to be avoided. A simple

illustration can be provided through considering the case of attempting to represent the volume and timing of water reaching the soil surface. There are numerous processes occurring here, such as interception by leaves, branches, shrubs and grasses, as well as evaporation and transpiration from each of these surfaces, throughfall, and stemflow. If the focus of a study is to partition interception into these various forms, it may be desirable to represent each of these processes separately (e.g. Rutter *et al.*, 1971), although this model is almost certainly over-parameterized. Where interception has been accounted for in a catchment-scale hydrological model, however, it is almost always satisfactorily represented through just one lumped "interception" parameter (e.g. Post, 1996, pp. 73–79).

3. Gathering the required input/output data at appropriate spatial and temporal scales Again, closely aligned with the data analysis focus, the idea here is to provide feedback to field data collection by creating a list of model data requirements. This list would be created based on an analysis of model over-parameterization to determine which parameters can be safely omitted or lumped. This may reduce the number of measurements required in the field, or it may require new measurements of the effect of a number of processes lumped together.

The longer term aim of the TDWG is to incorporate all of the knowledge that has been accumulated, making these incremental advances to then:

4. Create a new suite of hydrological models which are physically-based, yet parametrically efficient, and applicable to catchments where we have little or no hydrological data The examples provided above with regards to improvements in data collection, reductions in model over-parameterization, and parameterizing models in new ways which may exclude calibration, make it clear that currently incremental advances in model development are the way forward for the TDWG. However, there may well be a need to come up with a new approach, requiring modellers to go back to the drawing board and redesign their hydrological models from scratch.

Currently, surveying the hydrological literature, there appears to little support for such a radical step. It may not be necessary—it is possible that the incremental advances being proposed will allow a modified version of today's hydrological models to be used in areas where little or no hydrological data are available, i.e. for predictions to be made in ungauged basins.

If this is not the case, then it appears that the top-down approach will be the best way to achieve this goal. However, it is entirely possible that it will take some breakthrough in computing, programming, modelling, or data measurement and collection, to achieve the rather ambitious goal of designing new hydrological models which can be applied to ungauged basins.

WHERE TO FROM HERE?

If you have read this chapter and are interested in being involved in this exciting new area of research, then maybe you should consider joining the Top-Down Modelling Working Group (TDWG). Membership is open to anyone who feels they have something to contribute.

The inaugural meeting of the TDWG was held at the International Environmental Modelling and Software (iEMSs) Conference in Osnabruck, Germany, June 2004, and a follow-up meeting was held at the VIIth IAHS Scientific Assembly in Brazil, 4–9 April 2005. Future meetings are planned in conjunction with the Modelling and Simulation Society of Australia and New Zealand biennial Congress in Melbourne, Australia, 12–15 December 2005 (http://www.mssanz.org.au/modsim05/), at the Summit on Environmental Modelling and Software in Burlington, Vermont, USA, 9–12 July 2006 (http://www.iemss.org/iemss2006/), and at the IUGG XXIV General Assembly in Perugia, Italy 2–13 July 2007 (http://www.iugg2007perugia.it/).

To join the TDWG or to receive TDWG newsletters, please contact one of the authors of this chapter via the TDWG homepage (http://www.stars.net.au/tdwg/?welcome). Note that membership of this group is voluntary. Advances will only be made if the members of the group put the effort in. Your input would be warmly welcomed.

Contact details

David Post	david.post@csiro.au
Ian Littlewood	igl@ceh.ac.uk
Barry Croke	barry.croke@anu.edu.au

References

Alsdorf, D., Lettenmaier, D. & C. Vörösmarty (2003) NASA Surface Water Working Group: The need for global, satellite-based observations of terrestrial surface waters. *Eos Trans. AGU* **84**(29), 269

Boughton, W. C. (1984) A simple model for estimating the water yield of ungauged catchments. *Civil Engineering Trans.* **26**(2), 83–88.

Choubey, V. K. (1997) Monitoring turbidity with IRS-1A data. *Hydrol. Processes* **11**(15), 1907–1915.

Feng, X., Kirchner, J. W. & Neal, C. (2004). Spectral Analysis of Chemical Time Series from Long-Term Catchment Monitoring Studies: Hydrochemical Insights and Data Requirements. *Water, Air and Soil Pollution: Focus* **4**(2/3), 221–235.

Hreiche, A., Mezher, D., Bocquillon, C., Dezetter, A. Servat, E. & Najem, W. (2002). Parallel processing for a better understanding of equifinality in hydrological models. In: *Integrated Assessment and Decision Support* (ed. by A. E. Rizzoli & A. J. Jakeman) (Proc. First Biennial Meeting of the International Environmental Modelling and Software Society), vol. 1, 410–415. iEMSs, Manno, Switzerland.

Kirchner, J. W. (2004) The fine structure of water quality dynamics: the (high frequency) wave of the future. *Hydrol. Processes* **18**, 1353–1359.

Kirchner, J. W., Feng, X. & Neal, C. (2001) Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations. *J. Hydrol.* **254**, 82–101.

Littlewood, I. G., Croke, B. F. W., Jakeman, A. J. & Sivapalan, M. (2003) The role of 'top-down' modelling for Predictions in Ungauged Basins (PUB). *Hydrol. Processes* **17**, 1673–1679.

Post, D. A. (1996) Identification of relationships between catchment-scale hydrologic response and landscape attributes. PhD Thesis. Centre for Resource and Environmental Studies, Australian National University, Canberra, ACT 0200, Australia.

Post, D. A. & Jakeman, A. J. (1999) Predicting the daily streamflow of ungauged catchments in S. E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecol. Modelling* **123**, 91–104.

Post, D. A. & Jones, J. A. (2001) Hydrologic regimes of forested, mountainous, headwater basins in New Hampshire, North Carolina, Oregon, and Puerto Rico. *Adv. Water Resour.* **24**, 1195–1210.

Rutter, A. J., Kershaw, K. A., Robins, P. C. & Morton, A. J. (1971) A predictive model of rainfall interception in forests. I. Derivation of the model from observations in a plantation of Corsican pine. *Agric. Met.* **9**, 367–384.

Sivapalan, M, Blöschl, G., Zhang, L. & Vertessy, R. (2003a) Downward approach to hydrological prediction. *Hydrol. Processes* **17**, 2101–2111.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiodo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S. & Zehe, E. (2003b) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **48**(6), 857–880.

Sivapalan, M. & Young, P. C. (2005). Downward approach to hydrological model development. *Encyclopaedia for Hydrological Sciences.* Wiley, Chichester, UK (submitted).

Spate, J. M., Croke, B. F. W. & Jakeman, A. J. (2003) Data mining in hydrology. In: *MODSIM 2003 International Congress on Modelling and Simulation* (ed. by D. A. Post), vol. 1, 422–427. Modelling and Simulation Society of Australia and New Zealand, Canberra, Australia.

Wagener, T. & Wheater, H. S. (2002) A generic framework for the identification of parsimonious rainfall-runoff models. *Integrated Assessment and Decision Support* (ed. by A. E. Rizzoli & A. J. Jakeman) (Proc. First Biennial Meeting of the International Environmental Modelling and Software Society), vol. 1, 434–439. iEMSs, Manno, Switzerland.