5 Hydrological Simulation Modelling

STEWART W. FRANKS, STEFAN UHLENBROOK &
PIERRE ETCHEVERS

One of the corner-stones of modern hydrological science and practice is the use of simulation models. Whilst there are a large variety of modelling objectives, a whole range of different philosophical approaches also exist for achieving these different objectives.

The primary purpose of hydrological models is to make some kind of prediction. Due to the inherent variability of hydrological processes and only limited capabilities to measure them in the field, all models should be seen as approximations or simplifications of actual processes. Given the imperfect nature of our hydrological models, a key endeavour is to quantify the uncertainty associated with a model's prediction. In this regard, a vast array of tools has been developed and is at the disposal of the hydrologist.

In this chapter, key issues associated with simulation modelling in hydrology are reviewed. In the first section, the range of modelling objectives and philosophical approaches to their simulation is presented. The subsequent section addresses the key sources of uncertainty in modelling endeavours. Following this, different approaches to the estimation and propagation of uncertainty are assessed. The next section addresses the key issue of application and/or regionalization of hydrological models to ungauged basins. The chapter finishes with concluding remarks and a look forward to effective hydrological modelling for a range of hydrological objectives.

5.1 MODELLING OBJECTIVES

Hydrological simulation models are utilized in a seemingly vast array of environmental studies and investigations. Hydrology and the movement and behaviour of water are fundamental to many allied research areas. Consequently, different aspects of land surface hydrology may be modelled in quite different ways depending on the objectives of the individual study. A short, but not exhaustive, list of key predictive areas for hydrological simulation models includes the following:

- flood hydrology;
- drought hydrology;
- rainfall variability;
- evapotranspiration and land surface-atmosphere fluxes;
- erosion and sediments;
- assessing the impact of changed circumstances (e.g. land use/cover change, climate change);

– water quality and ecosystem health; and
– integrated (comprehensive) environmental assessment.

In many cases, the actual objective of a specific modelling activity may be directly related to the approach subsequently taken. For instance, a wide range of modelling approaches exists for simulating a basin's water yield (e.g. Singh, 1995). The hydrologist, if armed with rainfall–runoff data of sufficient quality and quantity, may choose a relatively simple transfer function-based hydrological model. However, if the study required assessing, for instance, the future effect of logging on the basin water yield, a more physically-realistic model containing explicit functions for the role of vegetation in the hydrological fluxes may be required. The transfer function approach, being largely a mathematical construct to mimic the observed behaviour, would struggle to represent future change in the basin as the derived model parameters might not easily be related to future changes in the basin's nature. This approach would, however, be robust for simulating its current behaviour as transfer function approaches are parsimonious and hence parameters tend to be well identified (e.g. Jakeman & Hornberger, 1993). In the case of a more distributed, physically-based approach, the model is distinctly more complex with more parameter values required to be either: (a) measured in the field, or (b) identified in a calibration procedure. Given the difficulties in both measuring appropriate parameter values *in situ* as well as the problem of obtaining non-unique sets of parameters in calibration, there will be significant uncertainty associated with the complex model (e.g. Beven, 1993).

Clearly, the choice of which type of model to use must therefore be governed by the questions to be answered by the model. If an extension of previous behaviour is required, a transfer function approach may be most robust. If simulating changes to the basin is the study objective, a more explicit and complex model should be used. The model needs to include an explicit representation (or a sufficiently accurate parameterization) of the physical processes that are assumed to govern the basin evolution. However, in common to both approaches in this example, significant uncertainty exists and hence the uncertainty of the modelling endeavour must be quantified.

5.2 DIFFERENT MODELLING PHILOSOPHIES

The example given in the previous section alludes to the existence of marked differences in modelling philosophies, all utilized within the wide field of hydrological science. At the extreme ends of a spectrum, these can be summarized as inductive *vs* deductive approaches. The inductive approach seeks to find and establish relationships between variables through the observation of the variables themselves. Inductive hydrological modelling approaches are consequently characterized by the data available defining the model in application to a specific study basin or area. The alternative modelling approach is the deductive approach. This approach follows logical reasoning in the construction of a model and may not have any recourse to the actual data/variable.

The inductive/deductive nexus is not unique to hydrology; they are two extremes of scientific approaches that often compete in any given scientific area despite the simple observation that they are in fact complementary and equally important. In the sphere of hydrological modelling, a good example of the diversity of approaches lies in rainfall–runoff modelling. Hydrologists tending toward the inductive approach may utilize data-based modelling approaches (e.g. Young, 2001) such as combinations of linear transfer functions, or highly parsimonious conceptualized rainfall–runoff models which are then calibrated to available data. Hydrologists tending towards the deductive approach may adopt a fine-scale distributed hydrological model with constituent processes being represented by laboratory-derived physical "laws" such as the Darcy-Richards equation (e.g. Abbott *et al.*, 1986). Another way of conceptualizing the two fundamentally different approaches to model building is the "top-down" *vs* the "bottom-up" approach (Sivapalan *et al.*, 2003a).

Strictly speaking, neither group is wholly inductive or deductive; the data-based models do have some structure that may represent how hydrological processes may be expected to exchange fluxes between stores, albeit not utilizing more complex laboratory-derived flux relationships. The more complex, distributed "physically-based" approaches still seek to evaluate the resultant model on available data and perhaps make subsequent adjustments on the basis of this comparison. Nonetheless, such philosophical differences in approach are a characteristic feature of modelling in modern hydrological science. The fact that alternative approaches to hydrological simulation exist is due to the significant uncertainty in the application of any model to any specific basin. In the following section, the sources of uncertainty in hydrological modelling are reviewed.

5.3 UNCERTAINTY IN HYDROLOGICAL MODELLING

There are three principally different sources of uncertainty in hydrological modelling: data uncertainty, model structure uncertainty and parameter uncertainty. Each source of uncertainty is discussed in more detail below.

5.3.1 Data uncertainty

Data uncertainty arises as a result of the imprecision or commensurability of the data used to force or else to calibrate a model. For example, in rainfall–runoff modelling, a model is forced with rainfall input data and the parameters are tuned until the model simulated outflow (runoff) corresponds best to the measured runoff from the basin area. Data uncertainty must be assumed to be present in both rainfall and runoff data.

Rainfall data may be corrupt for two key reasons. First, raingauges have measurement error associated with them. If two identical gauges are placed side by side, there will be an inherent, albeit hopefully relatively small discrepancy, between the two. For a single raingauge, errors may be induced by the local meteorology around the gauge (e.g. the wind effect or snowfall) and can be reduced by using a more sophisticated gauge design or installation procedure. Perhaps a far larger source of uncertainty associated with rainfall data lies in how representative a single gauge is of the total rainfall falling across some spatial extent within the basin area. This can be partially compensated for by the use of rain radars, which provide the rainfall distribution. (A more detailed discussion of different methods of rainfall measurement is provided in Section 4.2). Uncertainty associated with the runoff time series is arguably less problematic than with rainfall since, at least in principle, the runoff measured from a basin should be an integrated response from the basin area. However, runoff can be exported from the basin through unrepresentatively defined basin boundaries, and as underflow beneath the flow gauging structure.

Furthermore, flow gauging structures are required to be calibrated in the field at a range of flows. The subsequent rating curve is then only truly applicable for that range of flows. Over time, the channel reach upstream of the gauging structure may well change as a result of natural geomorphic/sedimentary processes, which may impart an error to the rating curve. Consequently, rating curves should be evaluated and checked on a regular basis. A larger source of error comes when the flow in the channel exceeds the range of flows used in the calibration of the gauging structure (observation of extremes). The estimation of flow outside this range represents an extrapolation of the rating curve. This may not be too critical so long as the flow remains within the flow gauging structure's geometry. When flows overtop the gauging structure, leading to complex channel geometry with potentially complicating shear patterns, peak flows can only be estimated after the flood through a subjective consideration of the flood inundation across the channel section (and potentially flood plain). This is somewhat unfortunate as it is often the high flows that are of most interest in many rainfall–runoff exercises and yet are the most uncertain. The observation of low flows is also often problematic, in many cases due to anthropogenic influences.

For more sophisticated models, data uncertainty also affects other terms of the hydrological balance, such as evapotranspiration or soil water storage. These data are very difficult to measure and new perspectives may be offered by remote sensing. However, the uncertainty of these new data sources is often not well known, as the comparison with ground-based observation is difficult (see Chapter 4, especially Sections 4.2–4.6 and 4.9 for detailed discussions of the use of remotely-sensed data of many hydrological parameters).

Another source of data uncertainty can be encountered when using meteorological data from a meteorological model (for instance, for an ungauged basin) or climate model (for a climate change impact study). The resolution of the input data is generally crude compared to the ground observation network or even to the basin surface area. Moreover, precipitation amount derived from atmospheric models can present biases for small time scales such as hours or days. Specific algorithms (e.g. anomalies methods, analogue methods, regional atmospheric models) have to be used to disaggregate the large-scale data and adapt them to the scale relevant for hydrological modelling. The uncertainty of the approach is very dependent on the method and the hydrological model and it can be estimated by testing the results for a well-instrumented basin.

5.3.2 Model structure uncertainty

Model structure uncertainty is the uncertainty associated with the choice of the hydrological model and its appropriateness with regard to the application at hand (e.g. Grayson *et al.*, 1992). Model structural error is arguably the hardest source of uncertainty to quantify; it is the error associated with not knowing *a priori* which processes and what process description best approximate the processes present at the application site. This uncertainty arises as a direct function of the limited observational capability of hydrology and the complexity of the processes in practice. Indeed, there are numerous hydrological models available that all purport to achieve the same predictive outcome (for instance, basin runoff) and yet comprise many alternative descriptions to achieve the same aim. Consequently, many conceptual models may have parameter values with more or less the same physical interpretation (e.g. root zone depth), but these values cannot be directly measured in the field or in the laboratory (because of monitoring difficulties or huge spatial variability). This problem is not limited to "conceptual" models. Even "physically-based" models may include parameterizations that differ from one model to another. The parameters are generally calibrated in these models, but, as the models contain different mathematical descriptions of how the processes are thought to operate (based on our incomplete understanding of hydrological process, see e.g. Section 7.1), when different models are calibrated to the same datasets very different parameter values may be returned, even in the absence of multiple optima in the response surfaces. This is due to the fact that parameter values are interlinked; in other words, the values of all parameters of a calibrated parameter set may influence, to some extent, the values of a given parameter.

Model structural uncertainty may also be present due to the non-stationarity of processes operating in the basin. For instance, significant changes to the hydrological flow pathways may occur over time, which may (in extreme cases) lead to significant changes in the behaviour of the basin. As all hydrological models tend to assume stationarity of the processes, it must be accepted that there is an inherent uncertainty associated with basin evolution, which is probably impossible to robustly detect and correct for in the model procedure.

5.3.3 Parameter uncertainty

Parameter uncertainty can be defined as the uncertainty associated with the specification of model parameters. If hydrological models were perfect representations of natural hydrological processes, then the parameters associated with our perfect model would be entirely physically meaningful. If techniques were available for quantifying the actual hydrological processes in the natural environment, then the resulting measurements would be input to our models as physically meaningful parameter values (for example, hydraulic conductivity, soil depth patterns).

In practice our models cannot reflect the vast complexity of each individual hydrological pathway and process. Moreover hydrological measurement techniques are imperfect. Whilst an error of an order of magnitude can be readily encountered in measuring hydraulic conductivity of a soil, it should be recognized that hydraulic conductivity is highly variable across a basin and with depth. The notion of measuring soil hydraulic conductivity at the scale of its variability would be prohibitively expensive as well as highly destructive to the basin itself!

Given the problems of spatial complexity, many hydrological models aggregate complex hydrological processes into simpler conceptual representations of their net effect. Such models have the distinct advantage of being relatively parsimonious with regard to the number of parameters that require specification. One disadvantage is that the resultant parameter values, whilst retaining a nomenclature that suggests a physical basis, are in fact at best "effective" parameter values or at worst entirely conceptual with hardly any physical basis.

An advantage of parsimonious conceptual models is that the relatively low number of parameters may be calibrated, given historic input and observed output data. Put simply, the model parameters are adjusted so that the model simulation best matches the observed hydrological response. The advantage of a model having fewer parameters lies in the tractability of the appropriate parameter values from the limited information contained within the provided input and observed response data. When too many parameters require calibration, there are often multiple combinations of parameter values that all yield similarly “good fits” in the calibration exercise.

5.4 MODEL PARAMETERIZATION

The early hopes of the pioneers of hydrological simulation models were founded on the advent of the ‘micro-processor’ which, for the first time, made computer-based simulation of hydrology feasible. These early days of hydrological modelling saw the development of sophisticated modelling schemes, representing each and every process the modeller could envisage. The early hopes were that model parameters that represent physical properties of the application basin could be measured in the field, plugged into the simulation model and accurate model simulations would follow. These efforts were confounded by the difficulties of achieving representative measurements of the constituent processes. Where processes are represented in a lumped manner, the spatial variability of hydrological processes precluded the derivation of “effective” measurements. In explicitly distributed models, even at the scale of the smallest distributed unit, spatial variability confounds the direct measurement of a representative parameter value. Other basin properties are very difficult to observe, particularly those relevant to subsurface flow pathways (although promising new techniques are being developed as discussed in Sections 4.4, 4.5 and 4.9). Consequently, modellers turned to calibration of hydrological simulation models to identify appropriate parameters.

5.4.1 Model calibration

Calibration is the act of adjusting model parameters and then comparing the simulated variable against observations of that variable. Many calibration strategies exist which differ in their approach to determining the “optimal” parameter value. The simplest form of calibration is manual tuning; in essence, the modeller alters certain parameter values according to some subjective judgements on how the model’s processes can be best adjusted to improve the comparison of simulated and observed variables. More sophisticated calibration schemes tend to be automatic: a calibration algorithm searches through the parameter space to identify the best fit between simulated and observed variables (e.g. Gupta & Sorooshian, 1985). The advantage of automated calibration routines is that very high dimensional parameter spaces can be efficiently explored. The argument for the continuing use of manual calibration is that, whilst inherently less efficient, the operator can guide the search to implicitly include expert notions of how that basin actually behaves, whereas automated routines may fit the observed data through unrealistic process parameterizations (Klemeš, 1986). Although automated routines are increasingly employed it is significant that the US National Weather Service still relies on manual tuning.

In either case, an objective function is used to quantify the comparison of simulated and observed variables. In many cases, the objective function is some kind of goodness-of-fit criterion, or else it is based on explicit assumptions about the errors (residuals). It has often been noted throughout the hydrological literature that the choice of objective function can have a marked consequence on the best-fit parameter set ultimately derived.

Regardless of the choice of objective function and the means employed for calibration, most conceptual basin models suffer from the same problems, namely the existence of multiple optima and the presence of high interaction or correlation between subsets of fitted model parameters (e.g. Soorooshian & Dracup, 1980). A consequence is that there may exist many parameter sets which span a wide range of the feasible parameter space, yet produce similarly high values of the objective function and even virtually indistinguishable simulated stream flow sequences.

The presence of multiple optima in the parameter space response surface is often due to the over-parameterization of models (e.g. Beven, 1993). Over-parameterization is largely a consequence of complex models permitting multiple alternative pathways for water fluxes to exit the basin. The data used to evaluate such models does not contain any information on these alternate pathways; consequently, the model has multiple combinations of parameters that can all reproduce the relatively uninformative observed flows. This has been referred to as the equifinality problem (Beven & Binley, 1992), meaning that different models or parameter sets yield equally good simulation results. This is particularly the case for distributed models and models with large numbers of parameters to be identified by calibration.

Such poor parameter identifiability may result in considerable uncertainty in the prediction of fluxes (e.g. Seibert *et al.*, 2000) and states used in model testing, and perhaps, more importantly, makes attempts to regionalize model parameters for the purpose of application to ungauged basins virtually impossible. Parameter identifiability is intimately linked to the information content of available data as well as to the complexity of the model. The modelling dilemma can be bluntly described as follows: a simple model cannot be relied upon to make meaningful extrapolative predictions, whereas a complex model may have the potential for prediction, but, due to information (data) constraints may be unable to realize it with little uncertainty.

5.4.2 Model verification

Ultimately, all hydrological simulation models should be tested for their ability to reproduce the dynamics of the basin system for which hydrological data are available. Testing hydrological models therefore requires data to achieve some objective measure of their performance, irrespective of the rationale for a given choice of model structure.

The typical way that this has been achieved in hydrological science has been through the use of the "split sample" or "calibration-validation" test (Klemeš, 1986). In this procedure a portion of the available data is set aside and not used in calibration, but is retained for the purpose of evaluating the calibrated model's simulation of an independent period of data.

<u>A Popperian view of hydrological model testing</u> Popper (1959) proposed a theory of falsification of testable hypotheses which is widely regarded as the basis of the scientific method. Both Dooge (1986) and Beven (1987) refer to Popper's theory as a basis for the critical assessment of hydrological models. Its key elements are elucidated below.

The scientific method provides an objective framework for the advancement of scientific knowledge. In our context it can be regarded as a model filter whose objective is to identify, in some sense, the best model(s). In the crudest terms, a scientifically-acceptable model filter involves iteration of the following two steps. First, the modeller conceptualizes a model structure or, more accurately, proposes a testable hypothesis. This is a creative act based on the modeller's world view which is much affected by current paradigms and knowledge, and by his/her experience. In the second step the creative act of the modeller is subjected to objective scrutiny in which the modeller attempts to falsify the hypothesis; this is what is referred to as model testing. The hypothesis must be falsifiable in the sense that it rules out the occurrence of certain events. Accordingly, the modeller compares model predictions against independently observed data to look for contradictions, a fundamental component of the scientific method.

In practice, all hydrological models can be easily falsified as the models themselves are often gross simplifications of reality which tend to lump complexity observed in the field at some scale. If, in any given application, the model underperforms in the representation of a particular aspect of the observed dynamic of the system, the models are then often adjusted, via a change in how a particular process is represented or else through a more judiciously selected parameterization. In the Popperian view of science, this can be seen as an auxiliary hypothesis: the model is adapted to fit the observations through additional complexity. However, the data utilized to reformulate the model cannot thereafter be used in a Popperian test of the model.

Popper argues that an auxiliary hypothesis should only be acceptable if it increases the degree of falsifiability of the model. Conversely, auxiliary hypotheses that add complexity to the model to protect it from falsification must be rejected. Popper's caution about auxiliary hypotheses protecting a model from falsification is another manifestation of the principle of parsimony. This principle requires the modeller to articulate the simplest model hypothesis consistent with the evidence/observations.

In the hydrological context the principle of parsimony does not appear to be widely articulated, possibly because reductionism coupled with a deep belief in the correctness of scaling up small-scale physically-based models has been the dominant paradigm. It has previously been argued that, if only streamflow data are available, simple models with four or five parameters based on a quick- and slow-flow conceptualization provide an adequate fit to the data (e.g. Jakeman & Hornberger, 1993). Moreover, the addition of more model structure and its associated conceptual parameters often leads to no significant improvement in fit yet introduces poorly identified parameters.

The principal weakness with this approach is that the model is typically tested against the same response (namely streamflow) to which it was calibrated. A split-sample test using streamflow data should be treated as the minimum requirement for testing model performance in an operational application. It has been observed that if the input data in the two split samples, one used for calibration and the other for testing, are qualitatively similar in the sense that the inputs span similar ranges, then the test lacks rigour. In essence, split-sample testing often represents a token challenge of the model hypothesis. Split-sample testing is obviously necessary, but is by no means a sufficient technique to challenge a model hypothesis.

The credibility of model testing depends on the power of the methods used to challenge or falsify the model hypothesis. The term "power" is a statistical concept describing the ability to discriminate between good and bad hypotheses. How the independent data are chosen can critically affect the power of the testing step. At the heart of the model testing problem is data, or rather the lack of it. With streamflow, typically the only observed basin response available, there is insufficient information to identify the conceptual parameters and, worse still, to meaningfully challenge the model hypothesis. One obvious and well-documented way to resolve this impasse, at least partially, is to couple more complex models with streamflow data augmented by other kinds of hydrological information relevant to the prediction task. We call this data augmentation, stressing that it is the information content of the data that is being augmented, not the data itself. Examples of data augmentation include streamflow and stream solute concentrations at different locations within the basin and other measurable hydrological fluxes or states such as soil moisture, piezometric levels and evapotranspiration at selected locations (e.g. Ambroise *et al.*, 1995; Mroczkowski *et al.*, 1997; Franks *et al.*, 1998; Uhlenbrook *et al.*, 2004). Such data present additional opportunities to test or falsify model hypotheses but often require additional model complexity via auxiliary assumptions and parameters.

5.5 INTERNAL STATES AND THE ISSUE OF COMMENSURABILITY

Much recent research has focused on seeking additional measures of hydrological states and fluxes with the aim of providing more information for the calibration and testing of hydrological models. Examples include:

– fluxes (runoff, evaporation) measured at small scales;
– borehole logs to characterise water table dynamics;
– distributed soil moisture fields;
– hydrochemical data and environmental isotopes, and
– measures of contributing areas.

In many cases problems are encountered using these additional data. Model structures are grossly simplified representations of reality and field-measured fluxes may not be directly comparable to the quantities predicted by the models. This issue of commensurability of data and models can often mean that auxiliary hypotheses (in other words, additional model complexity) need to be introduced if the two are to be compared. These additional degrees of freedom have the effect of reducing the information content of the data and reducing the opportunity for falsification. This is not to say that the additional data cannot aid the robust testing of hydrological models, rather that their worth is not immediately apparent and that the incorporation of such information needs to recognize this problem. In the following sections the issue of commensurability is explored using specific examples.

5.5.1 Examples of commensurability

<u>Local measures of water levels</u> Within the TOPMODEL framework (Beven *et al.*, 1995), the subsurface discharge at a given point in the landscape where there is surface saturation is (equation 5.1):

$$q_i = T_i \tan\beta \tag{5.1}$$

where q_i is the local subsurface discharge at a point i (m^2 s^{-1}), T_i is the transmissivity (m^2 s^{-1}), and $\tan\beta$ is the local slope. Transmissivity is usually parameterized within TOPMODEL as a lumped effective basin-scale parameter. This is appropriate as distributed information about local transmissivities is usually unavailable.

To utilize distributed point measurements of local water table dynamics, either as additional calibration data or as internal testing of model dynamics, the effects of local deviations from the basin effective transmissivity must be acknowledged and permitted. An interpretive model (or auxiliary hypothesis) must be specified to permit local deviations of transmissivity and porosity according to the following (equations 5.2 and 5.3):

$$q_i = T_0 \tan\beta \exp\left(\frac{-D_i}{m}\right) = T_0 \tan\beta \exp\left(\frac{-z_i \Delta\theta_i}{m}\right) = T_i \tan\beta \tag{5.2}$$

$$\log_e\left(\frac{T_i}{T_0}\right) = \frac{D_i}{m} = \frac{z_i \Delta\theta_i}{m} \tag{5.3}$$

where T_0 is the basin effective transmissivity at saturation (m^2 s^{-1}), m is the exponential transmissivity decline parameter (m), z_i is the local depth to the water table (m), D_i is the storage deficit at point i (m), and $\Delta\theta_i$ is the effective water content change per unit depth in the unsaturated zone (porosity) (–).

To calculate the local corrections to transmissivity and porosity at the site of a borehole log, equation (5.3) must be applied to two water table observations. In essence, the water table observations are made commensurate by assuming that both T_i and $\Delta\theta_i$ are constant over a range of water table depths, and using these two local parameters to scale the observations in line with the predicted storage deficits. It is therefore clear that the interpretive model necessitates additional degrees of freedom in the model structure (i.e. the parameters representing the local deviations of transmissivity and porosity) to enable comparisons of local point measures against the model storage deficits. These additional degrees of freedom offer, on the one hand, the possibility of using additional data in model calibration and testing, but on the other hand, it has the effect of reducing the informative content of the data.

Recent applications of water table measures to constrain the predictions of hydrological models have revealed some utility of this approach (e.g. Lamb *et al.*, 1997, 1998). In their study of the HBV* model, Seibert *et al.* (1997) found that no single parameter set could reproduce both discharge and water table data necessitating the rejection of the model as a hypothesis. They subsequently modified the model structure to enable consistent reproduction of both variables. However, while distributed water table depths are intuitively an obvious choice of additional measurement if planning a field campaign, their local nature and the necessity of an interpretive model as described above, reduces the power to falsify model hypotheses.

<u>Microwave remote sensing</u> Research in the microwave remote sensing domain has offered some promise of delivering distributed datasets of soil moisture useful for model calibration and testing. However, several problems exist. The accuracy of

* Named after the former Hydrologiska Byråns Vattenbalansavdelning (Hydrological Bureau Waterbalance Section), at SMHI, the Swedish Meteorological and Hydrological Institute, where the model was originally developed.

retrieved absolute soil moistures is complicated by factors such as differences in soil characteristics, surface roughness (especially vegetation type and density), and topography (see for example, Lin *et al.*, 1994; Dubois *et al.*, 1995; Famiglietti *et al.*, 1999). Even in ideal conditions, microwave remote sensing can detect soil moisture only in the upper few centimetres of the soil, yet both runoff generation and evapotranspiration may be more strongly controlled by deeper layers. Further limitations result from the reduced sensitivity of the radar signal to soil moisture changes close to saturation and the specular nature of the signal in areas where water becomes ponded leading to decreases in the backscattering coefficient (Bertuzzi *et al.*, 1992).

The problems of deriving surface roughness for microwave remote sensing are significant. It is also apparent that even if robust measurements could be achieved, the value of surface measures, as opposed to profile (depth-integrated) measures, remains questionable.

<u>Saturated areas</u> It was shown earlier that incorporating borehole water table dynamics may not be straightforward due to the requirement to adapt a given model structure to account for highly localized basin characteristics unparameterized in lumped conceptual models. The worth of highly localized measures of hydrological fluxes must always be questionable. An alternative data augmentation approach would seek more integrated measures of basin response, for example the areal extent of surface saturation or contributing areas. This is intuitively appealing as the effects of real-world high-frequency local variability may be integrated into a measure that is more commensurate with simulated responses.

Such an approach was utilized by Gineste *et al.* (1998) who analysed microwave backscatter behaviour in a 1.2 km^2 subcatchment of the Naizin basin, northern France, where saturated areas were observed every three days concomitantly with the acquisition of satellite data. The areal extent of the saturated area could not be assessed for each individual date but a saturation potential index (SPI) was derived on a pixel by pixel basis from the temporal standard deviations for each pixel computed on the filtered images of backscatter. The rationale for such an approach was that areas which would be most prone to saturation would be those that exhibited the least temporal variability of moisture (due perhaps to significant lateral redistribution of upslope moisture). As noted above, an active microwave signal is affected by both soil moisture content and soil roughness. By using a temporal change approach the obfuscatory effect of roughness, assumed to be constant in time, could be negated.

Utilizing an approach based upon combining threshold values of the topographic index and the SPI, the limited ground-truth data available were extrapolated to yield estimates of the total basin saturated area extent (Franks *et al.*, 1998). As this extrapolation is inherently uncertain, multiple combinations of the topographic index and SPI were used and a range of feasible saturated areas produced. These derived estimates of saturated area fraction were then used as a secondary modelling objective against which model simulations were compared and rejected. The saturated area information, though noisy, significantly reduced uncertainty in the model parameters, particularly the basin effective saturated transmissivity parameter. Perhaps of more significance, the uncertainty in the model predicted discharges was markedly reduced; the uncertainty of the peak discharge was reduced to approximately 30% after inclusion of the saturated area data.

5.5.2 *Commensurability and interpretative models*

The three examples above all provide opportunities to further discriminate between different models and/or parameterizations. Additional information may be utilized to provide further conditioning or testing of the model, but the inclusion of additional information is not straightforward as the data are not necessarily commensurable with the model. Direct insertion of the additional data is precluded by the fact that it does not necessarily correspond directly to a model simulated flux or state. Interpretative models are required to transform the data in some way to be commensurable with the model. This inherently requires additional model complexity and weakens the apparent informative content of the data. Nonetheless, the inclusion of such additional data can still provide further information with which to refine or verify models. It also offers the possibility to predict internal stages (e.g. water table fluctuations) with more confidence after such data are included in the model calibration procedure. This is often needed in model studies that are part of an extensive environmental assessment, for instance.

5.6 *UNCERTAINTY ESTIMATION*

Hydrological models simulate the dynamics of some or a number of aspects of terrestrial biophysical behaviour. Common to all these models are the following features:

– They are idealizations/simplifications of complex natural systems which have both spatial and temporal dynamics.
– Forcing inputs such as climate are uncontrollable, meaning controlled experimental designs are difficult to implement.
– Observations of forcing inputs and system responses are subject to (often considerable) error, on account of the spatial extent of the system.

These features conspire to introduce a complex and substantial uncertainty into any endeavour to identify, calibrate and validate a hydrological model and to use it for the purpose of prediction.

A common requirement of hydrological models is the ability to extrapolate model predictions into system dynamics that have not previously been encountered or used for model calibration exercises ("predictions in the unknown"). This permits the use of hydrological models as "experimental laboratories" in which the response of the system to external perturbations, for instance, significant land use change, or climate change, can be assessed. Typically, physically-based and/or conceptual models are built for this purpose. These models contain mathematical descriptions of individual processes in the hope that the constituent individual process descriptions can be sufficiently identified by measurement of field parameters. When this is not possible, calibration of the model is performed with the same aim, namely to adequately identify the individual process descriptions. When the model is applied to an extrapolative forcing circumstance (such as land use change), it is assumed that the physical basis underlying the individual process descriptions will provide the most representative prediction of the natural system.

A key problem with this current paradigm is that many physical processes, of which not all are well understood (e.g. subsurface stormflow generation), may influence the integrated response of the natural system at hand. The fact that complex, non linear multiple processes interact to produce a variable of interest means that a calibration exercise performed against that variable is a non-unique "inverse problem", characterized by multiple parameter sets that can simulate the calibration variable equally well. Whilst many *a priori* parameter sets may be rejected through comparison to calibration data, many remain as acceptable simulators of the system. The extreme degree of this problem renders modelling studies in all areas of hydrological research subject to unacceptable uncertainty with respect to predicting system responses to change. This uncertainty is poorly understood, indeed misinterpreted, primarily because our current approaches for dealing with model identification and calibration make simplistic assumptions about errors in environmental models.

5.6.1 Current approaches to uncertainty estimation

Three schemes, presented in the hydrological literature, typify current approaches to estimating the prediction uncertainty of models. Firstly, Beven & Binley (1992) adopted an informal Bayesian approach called GLUE whereby additional information can be incorporated and used to refine the model parameter inferences (e.g. Beven & Freer, 2001; Uhlenbrook & Siebert, 2005). A problem with this approach is the degree of subjectivity in assessing the relative acceptability of competing parameter sets (i.e. subjectivity in the definition of the likelihood function). Moreover, all sources of uncertainty are treated through parameter uncertainty simultaneously and the method is very computing intensive, often requiring several thousands of model runs[†].

Sorooshian & Dracup (1980) developed criteria for parameter optimization which were guided by the statistical behaviour of model residuals. However, errors in input and response and model structure were lumped together. The NLFIT scheme (Mroczkowski *et al.*, 1997) implements a stricter and more traditional Bayesian implementation than GLUE using the Sorooshian & Dracup error approach. Nevertheless, it has been found that the error probability models used in NLFIT are unsatisfactory when used with models that make predictions over short intervals. For example, errors in simulating monthly streamflow are adequately described by traditional error models, whereas errors in hourly streamflow cannot be satisfactorily described using commonly used error models of heteroscedasticity and persistence as the real error structure is much more complex. In a third type of scheme, Yapo *et al.* (1998) developed a multi-objective function approach whereby many different objective measures of fit are used to derive the Pareto-optimal set of parameters. Like GLUE, this scheme suffers from subjectivity in the selection of the number and type of objective functions.

One common feature of all these calibration schemes is the lack of any treatment of errors in the forcing data, such as rainfall, potential evapotranspiration, and nutrient inputs. At first sight this may be surprising given that the significance of such errors was appreciated decades ago. In 1967, Fiering showed that, in the presence of storage, errors in estimating rainfall input will lead to errors in rainfall-dependent model outputs such as streamflow which persist for a considerable time after the input has occurred as

† As far as the authors are aware, the record for the largest number of Monte Carlo runs is held by Iorgulescu *et al.* (2005) who carried out more than two billion model runs for a combined hydrological-hydrochemical modelling effort.

the storage maintains a "memory" of the input error. Such errors occur in virtually all hydrological models because temporary storage is virtually a universal feature in natural systems. Often, long runs of persistent under- or over-estimation are apparent in model simulations, arising from input error and its persistence. Even if the streamflow was measured with negligible error and the rainfall–runoff model used was structurally correct, errors in rainfall input can induce persistent errors. If these errors are incorrectly attributed to model error, the credibility of the model is unfairly undermined and the accuracy of predictions is unduly devalued. All models are simplifications at some level and hence must be in error. Processes that are known to exist and/or dominate in the field may be described in a number of ways with varying degrees of complexity. Uncertainty will therefore arise from the range of possible descriptions that may be employed to describe the same phenomena, as evidenced by the large number of available hydrological models published within the scientific literature.

The statistical difficulties in dealing with input forcing errors are formidable and, as a result, little research has been undertaken to tackle this problem. The few exceptions have not proved to be successful. For example, Kitanidis & Bras (1980) employed the Kalman filter, a conceptualization ideally suited to distinguishing between model response and input error provided the model can be suitably linearized and errors follow a Gaussian distribution. Unfortunately the promise of the Kalman filter has never been realized, largely because of the very significant nonlinearity of environmental models (Duan *et al.*, 1992) and the need to linearize the hydrological model with regard to parameters and inputs. The difficulty is that the assumptions of traditional likelihood/optimization theory (namely: a correct model structure, Gaussian error distribution, independence amongst measured variables, and independence of temporal errors) are violated in hydrological modelling. Given that hydrological models are excessively simple representations of reality, a correct model structure cannot be assumed. Moreover, it is well known that measures of pertinent variables, such as rainfall, discharge and evaporation, contain significantly nonlinear errors. To date, an appropriate theory for distinguishing between model, response and input errors has not been defined.

Significant model uncertainty may also exist due to the omission of a key control on the response of the natural system. Numerous mechanisms exist by which moisture may be transported to the basin outlet, each of which is uncertainly quantified in the field. However, recent attention to the role of subsurface preferential pathways has indicated that, despite their limited volume, they can in fact dominate the hydrological response of certain types of basins. The problem in representing such pathways lies in the fact that they are often impossible to observe. In fact, studies demonstrating the effects of such pathways are largely comprised of indirect evidence (e.g. isotope studies of old/new water contributions to basin outflows). It is therefore clear that significant uncertainty must be associated with the choice of model structure in any application attempting to simulate basins where subsurface flow pathways dominate the basin response.

To differentiate the role of model structure error from that of error in the forcing data and parameter uncertainty, the relative roles of parameter uncertainty and input forcing errors must be quantified. If, after permitting parameter uncertainty and input error, errors remain in the reproduction of the model outputs, then this residual error must be representative of model structure error.

5.6.2 Alternative approaches to uncertainty estimation

As indicated earlier, typical applications require models to be calibrated and validated against observed flows given meteorological input data. Numerous schemes have been developed to address this task, including MOCOM (Gupta *et al.*, 1998), GLUE (Beven & Binley, 1992; Franks *et al.*, 1998) and NLFIT (Mroczkowski *et al.*, 1997). However, Franks & Kuczera (2002) showed that all these contemporary approaches oversimplify the uncertainty associated with the appropriateness of the model to the study area and the measurement of both meteorological forcing and runoff. By ignoring these key sources of error and uncertainty, typical calibration schemes have been shown to produce biased estimates of parameters which may not have any real physical meaning (Beven, 1989).

In contrast, the Bayesian Total Error Analysis (BATEA) methodology (Kavetski *et al.*, 2003) is based on a "total-error" approach using rigorous Bayesian methods. The BATEA scheme has been tested against both real and synthetic data to verify the procedure (Kavetski *et al.*, 2003) and results to date indicate that the methodology is robust and yields parameter inferences that are in principle less biased due to potential input errors. To illustrate the degree of bias that arises as a function of ignoring rainfall forcing error, Kavetski *et al.* (2003) examined the probability distributions of a key model parameter following traditional standard least-squares (SLS) identification and the BATEA methodology. The resultant posterior parameter distributions were very different, but it was noted that tight constraint of uncertainty following the traditional SLS approach indicated a wholly false confidence in the input error-biased parameter distribution. It is therefore clear that ignoring key aspects of data uncertainty results in marked parameter bias. BATEA is unique in that it permits the inclusion of error models for observations (both input and output) within any hydrological model. It provides a rigorous framework for dealing with data, parameter and model uncertainty. Importantly it has revealed the key role of rainfall data error which, if ignored, will produce biased parameters and hinder attempts at regionalization. The direct consequences of largely ignoring error in input rainfall data are: less robust parameter inferences (estimates of appropriate model parameters), biased optimum parameter sets, and incorrect estimates of predictive uncertainty as a function of the above. The indirect consequences are: model parameters cannot be inverted for use as model-scale estimates of meaningful quantities and model predictions of "future" events (e.g. response to climate change, land use change) will be in error. Indeed, the validity of all model predictions may be inappropriately open to question, through failures of model predictions.

Throughout many areas of science and engineering, researchers have proposed, constructed and developed models for the purpose of understanding, simulating and predicting the behaviour of many natural systems. However, often the apparent success of such model applications, and hence their subsequent employment for a range of research purposes, is not fully tested as little is presented with regard to the uncertainty of the model's subsequent predictions. This apparent success, coupled with appealing model sophistication, may be used to justify inaccurate and largely untested inferences on the basis of these modelling studies. There is a critical need to improve the realism of parameter-model inference techniques to achieve more realistic predictions.

Given the problems associated with model and parameter identification and the inadequacy of current uncertainty estimation frameworks to account for forcing errors, more holistic treatment of the sources of uncertainty is required. The successful development and implementation of such a methodology would provide a new and more realistic methodology for the assessment of predictive uncertainty for application throughout the fields of environmental science and engineering. If this was achieved the following outcomes might be expected:

– quantification of the relative roles of the key sources of uncertainty;
– identification of a suitable model structure;
– refined predictive uncertainty estimation;
– refined parameter uncertainty estimation;
– refined forcing error identification;
– increased insight into the minimization of error and uncertainty;
– increased ability to test model hypotheses; and
– improved knowledge to design targeted field programmes (measuring the data that are really needed to reduce the uncertainty of our models).

5.7 APPLICATION TO UNGAUGED BASINS

In recent years, the importance of developing globally consistent hydrological modelling and parameter estimation techniques has become apparent. In 2001, IAHS declared a major initiative, the International Decade of Ungauged Basins (Sivapalan *et al.*, 2003b). Similarly the Global Energy and Water Experiment (GEWEX) has recently created MOPEX (Model Parameter Estimation Experiment, Schaake *et al.*, 2001). The aim of these initiatives is to derive representative models and their associated parameter values for both gauged and ungauged basins across the globe in order to provide more accurate tools for water resources management in ungauged basins.

To attain the full utility of hydrological modelling, hydrological models must be applicable to *all* basins. In applying models to *ungauged* basins where no records are available with which to calibrate or verify the model, significant uncertainty is associated with the *a priori* specification of both model and parameters. The applied basin model is assumed to correctly represent the dominant flow generating mechanisms, whilst its specified parameters are assumed to be meaningful.

To identify common physical characteristics of basins that may then relate to model parameters, regionalization approaches have been developed (for example, Post & Jakeman, 1996; Abdulla & Lettenmaier, 1997; Sefton & Howarth, 1998; Seibert, 1999; Yokoo *et al.*, 2001). Typically, a single conceptual hydrological model is selected and subsequently calibrated to the records of a number of gauged basins. Regression analysis is then performed to assess the consistency of individual parameter values (for instance, hydraulic conductivity, root zone store) against measured physical basin characteristics (e.g. soil texture, vegetation). These approaches have revealed some success in finding significant relationships for approximately half of the model parameters, but the predictive capability of models varied (e.g. Post & Jakeman, 1996; Seibert, 1999). Apparent parameter consistency or variability between different basin characteristics and responses in previous studies will have been significantly biased due to: (a) the assumed single applicable model; and (b) the calibration scheme employed providing biased parameter estimates. As noted by Post & Jakeman (1996), improvements in predictive capability would be attained through better understanding of the controls on hydrological response.

The problems of natural variability and data scarcity have meant that the development of a single hydrological model based upon on a fundamental "physics" of hydrology is thought unattainable (Beven, 1989, 1993). As a consequence, hydrological models are largely "conceptual", in that they constitute simplified representations of the mechanisms perceived to dominate the hydrological problem at hand. This means that there exists a whole range of different hydrological models to achieve particular tasks at specific spatial and temporal scales.

5.8 CONCLUDING REMARKS AND FUTURE CHALLENGES

Much hydrological science has focused on and sought fundamental physical insights into the behaviour of water in the natural environment. These fundamental insights have, however, been gained typically at small scales. For instance, major successes in fundamental hydrological sciences include Darcy's law, Richards equation, the physics of evapotranspiration, the physics of sediment entrainment and transport and hydrogeochemical interactions. However, much of modern hydrological science has also largely been based on the development and application of integrated models, which include the classical hydrological components and also modules from neighbouring disciplines, such as ecology and socio-economics and which are typically developed and utilized at much larger scales. Finally, another major section of hydrological science has focused on the development of complex measurement techniques for estimating, for instance, evapotranspiration (e.g. eddy covariance, Bowen ratio) or for making large-scale hydrological observations through remote sensing (e.g. soil moisture, evapotranspiration, terrain).

Despite the major efforts in each of these sub-areas of hydrological science, there is an apparent disconnect between them. It is often difficult to integrate the diverse achievements attained in the fundamental principles (or physics) of hydrology, practical hydrological modelling, and novel hydrological measurement techniques. As a consequence, much research focuses on model identification as practical models cannot be easily parameterized with field properties. There are however many different approaches that will assist in achieving integration within hydrological science. Developed calibration approaches include SLS, SCE, GLUE and Bayesian methods. All are entirely reliant on observed data and lead to the necessity of "parsimonious" models, given limited data and the limited information content of those data. The problem of ungauged basins is highly significant, as by definition no calibration data are available and predictions rely on transferring insights either from nearby basins or through regionalization.

As noted in the previous sections, there are many problems associated with regionalization. These include the assumptions that the runoff model itself is applicable everywhere, that calibration schemes need only focus on fitting to discharge data alone, and that forcing data are perfect. These assumptions invalidate any real meaning in derived model parameters so that it is difficult to justify retrieved parameter values and hence regionalization relationships are not robustly achieved. In many cases, different models/approaches are utilized for regionalizing different aspects of the hydrological regime; for example, flood *vs* low flow regionalization techniques.

Arguably, we have reached the limit of classical regionalization capability using current calibration techniques and standard rainfall–runoff data alone. There is a clear need to evaluate and integrate alternative methods/techniques as well as to augment model applications with alternative data. These are the key challenges to be addressed if meaningful progress is to be made in hydrological modelling, particularly in the pursuit of making predictions in ungauged basins.