INVITED COMMENTARY

HP TODAY

HYDROLOGICAL PROCESSES
Hydrol. Process. **25**, 661–670 (2011)
Published online 16 November 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hyp.7899

Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing

Dmitri Kavetski[1]* and
Martyn P. Clark[2]

[1] *Environmental Engineering, University of Newcastle, Callaghan, NSW, Australia*
[2] *Research Applications Laboratory, National Center for Atmospheric Research (NCAR), Boulder, CO, USA*

*Correspondence to:
Dmitri Kavetski, Environmental Engineering, University of Newcastle, Callaghan, NSW, Australia.
E-mail:
dmitri.kavetski@newcastle.edu.au

Received 12 September 2010
Accepted 15 September 2010

Why Worry about Numerics Given so Many Other Problems?

Hydrologists often face sources of uncertainty that dwarf those normally encountered in many engineering and scientific disciplines. While a structural engineer designing a wall of a building can subject multiple bricks to repeated strength tests and simulate the full non-linear behaviour of individual bricks, joints and reinforcing bars using finite-element models applied at the scale of millimetres, we as hydrologists often represent highly heterogeneous catchment systems, which may include complex stream networks, preferential flowpaths, varied vegetation, land use and geology, using highly conceptualized lumped models. Moreover, we often force these models with rainfall data from a single, daily recording gauge well outside of the catchment. Given the simplicity of our models, does it really matter how they are implemented?

It is then perhaps unsurprising that when asked 10 years ago whether the popular TOPMODEL (Beven, 1997) calculates its fluxes based on its storage at the start or at the end of a time step, a colleague responded with a largely indifferent shrug—'does it really matter, given all other errors we are making?' Similarly, another colleague often referred to mysterious 'pits' in the objective function surface of a hydrological model, which prevented the convergence of a standard parameter optimization code we got off a floppy disk in a numerical analysis book. What is causing those?

It would hardly be an exaggeration to say that objective function difficulties have caused hydrologists many headaches over the last 40 years. For example, Hendrickson *et al.* (1988) demonstrated that Newton-type parameter optimization terminated all over the objective function space depending on the initial search point. In a seminal paper, Duan *et al.* (1992) showed glacial-like objective function landscapes, with sharp spikes and otherwise messy geometry, and designed a global evolutionary method (the Shuffled Complex Evolution search, or SCE) that was better suited to such problems. Around the same time, Beven and Binley (1992) proposed the generalized likelihood uncertainty estimation (GLUE) framework, which was at least partially motivated by problems associated with searching for parameter optima of complex model response surfaces. Indeed, the dotty plots reported in GLUE studies often show significant distortions and irregularities. Given the undeniable empirical evidence of objective function complexity, the theme was picked up in the design of Monte Carlo Markov Chain (MCMC) algorithms, with techniques such as Shuffled Complex Evolution Metropolis (SCEM) and DiffeRential Evolution Adaptive Metropolis (DREAM) (Vrugt *et al.*, 2003, 2009) aiming to improve the sampling of geometrically complex and multi-optimal parameter distributions.

Quite remarkably, in many cases, model analysis complexities are consequences of numerical artefacts in the model implementation—often, literally, depending on whether the start-of-step or end-of-step fluxes are used (Clark and Kavetski, 2010)! The extent of problems can be

staggering, including (i) degraded performance of parameter optimization and uncertainty analysis algorithms, (ii) erroneous and/or misleading conclusions of sensitivity analysis, parameter inference and model interpretations and, finally, (iii) poor reliability of a calibrated model in predictive applications. Indeed, at times model behaviour becomes patently absurd (see Section on Harbingers of Problems).

How can this be? While various explanations could be advanced, we note that hydrologists have, quite correctly, attributed objective function complexities to model non-linearities (and large data errors). Yet, the hydrological community has often lacked a willingness to thoroughly question 'what, more precisely, is the origin of these non-linearities?' and 'is there something that is aggravating the non-linear behaviour and resulting problems?' Instead, we have often launched straight into inventing new calibration algorithms and paradigms to handle these complexities, rather than seeing if these problems can be reduced, or even avoided, in the first place.

This commentary surveys and discusses recent work on the impact of poor numerical techniques in conceptual hydrological models, relating it to the important questions of reliable model development and the hypothesis-oriented paradigm in hydrological sciences. In particular, we excoriate the frequent disregard of numerical errors by hydrologists who tacitly assume them to be insignificant and dominated by data and structural uncertainty. Such perception is almost akin to arguing that an approximate problem might as well be solved using a broken calculator, and, as illustrated here, may have caused hydrologists decades of unnecessary headaches.

Numerical Troubles

Mathematical structure of hydrological models

Revealing insights into the nature of model non-linearities can be gleaned from inspecting the formulation of hydrological models. For example, the continuous-time state-space form of an exponential store is

$$\frac{dS(t)}{dt} = P(t) - Q(S(t)) \tag{1a}$$

$$Q = b\exp(kS) \tag{1b}$$

where $S(t)$, $P(t)$ and $Q(S(t))$ are, respectively, the storage, inflow and outflow at time t, $\{k, b\}$ are (positive) model parameters, and $S(t_0) = S_0$ is the initial condition. Note that in Equation (1), the storage is defined relative to an implied datum (i.e., $S = 0$ has no special significance) and is unconstrained (i.e., $S \in (-\infty, +\infty)$) (e.g. Michel *et al*., 2003).

More realistic models will contain additional components representing surface runoff, evapotranspiration and/or other processes; multiple coupled equations will be used to describe water balance in distinct storage elements (e.g. canopy, snow and soil) and, in the case of distributed models, multiple state variables will be used to represent spatial variability (Cherkauer *et al*., 2003; Ivanov *et al*., 2004, and others).

In practice, regardless of the model structure, many hydrologists write the model algorithm and computer code directly from their perceptual understanding, bypassing the formulation in continuous-time state-space form. Many model codes, and even descriptions of models (see the listings in Singh and Frevert, 2002a), are styled as follows:

```
Srz = Srz_ini ! initialize store
DO i = 1, n   ! loop over time steps
  ! calculate outflow using parameters
  outflow(i) = b * exp(k * Srz)
  ! update storage
  Srz = Srz + inflow(i) - outflow(i)
END DO
```

This can be recognized as the explicit Euler approximation of Equation (1) over a series of discrete times t_n with spacing Δt,

$$S_{n+1}^{(\mathrm{EE})} = S_n + \Delta t P_n - \Delta t Q(S_n) \tag{2}$$

and is based on the fluxes at the beginning of the step (Press *et al*., 1992; Butcher, 2008). Except in the limit $\Delta t \to 0$, even a single step of an approximation such as Equation (2) incurs local errors that depend not only on the step size but also on the non-linearity of the solution and its derivatives. Moreover, because an entire simulation time period comprises many time steps, global errors will arise from the accumulation of local errors.

Also, note that the code snippet above reflects the style of many current hydrological models, which implicitly embed the time step size into their equations (in the above example, the time units of parameter b and the time units of the forcing data become intrinsically linked to the time step size used in the model simulation). This situation appears quite common. For example, in the popular GR4J model, the step size is not stated in any of the model equations (Perrin *et al*., 2003): the tacit assumption $\Delta t = 1$ (days) is 'hardcoded' directly into the model and its software. This introduces time scale dependencies beyond those present on the underlying state-space model, and makes it unnecessarily difficult to apply the model to data with a different temporal resolution.

Harbingers of problems

Hydrologists have often observed and reported numerically poor behaviour in some specific models. For example, Michel *et al*. (2003) noted that common implementations of the exponential store can produce absurd behaviour. Consider the explicit Euler approximation $S_{n+1}^{(EE)} = S_n + \Delta t[P_n - b\exp(kS_n)]$ applied in two near-identical cases: both cases use the same parameter values $\{k, b\}$ and receive the same precipitation input P_n, but case A has a higher initial storage S_n than case B. Because of the exponential dependence of the outflow on the storage, there is always a point beyond which increasing the storage S_n at the start of the step will result in decreasing the storage estimate $S_{n+1}^{(EE)}$ at the end of the step. Thus, case A could end up drier than case B despite having higher initial storage (with all parameters and forcing being identical). Such behaviour makes no hydrological or mathematical sense and never occurs in the exact solution of Equation (1).

Related problems can also arise with respect to increases in the rainfall P_n. For example, if the rainfall is added to the store *before* computing its outflow [as recommended by Michel *et al.* (2003); this could be viewed as an 'operator-splitting' approximation technique, which allows applying fluxes in various distinct sequences (Schoups *et al.*, 2010)], the store can become increasingly drier by the end of a step as its rainfall input is increased. Note that adding the rainfall at the end of the time step can simply postpone the absurdity till the next step. In either case, a 'sawtooth' pattern illustrated in Figure 1 can develop. Unsurprisingly, permitting such behaviour can lead to catastrophic errors in model predictions (see Figures 4 and 6 of Michel *et al.*, 2003).

These are not isolated instances. Michel *et al.* (2005) subsequently reported that the traditional implementation of the Soil Conservation Service Curve Number (SCS-CN) approach to soil moisture accounting widely used in rainfall–runoff modelling also suffers from mathematical inconsistencies, again with respect to initial conditions. Although Michel *et al.* (2003, 2005) mended the problems in both the exponential and SCS-CN models by employing analytical integration techniques, as we elaborate in a later section, analytical solutions are seldom applicable to general hydrological and environmental models, especially in spatially distributed contexts.

Can model complexity protect us against poor numerics?

A sceptical hydrologist will correctly note that few, if any, 'real' hydrological models are as simple as Equation (1). Could it be that additional model complexity, e.g. multiple flow pathways, ameliorates these defects? Can problems arising in simple test cases with known exact solutions go away in more complex modelling scenarios?

Figure 2 shows the behaviour of a more realistic model structure, where the non-linear exponential store is supplemented with a surface runoff term (based on a topographic index with a Gamma areal distribution, Sivapalan *et al.*, 1987). Model simulations used parameter values within the ranges reported for similar TOPMODEL-type formulations (Beven, 1997). Consider the case of constant precipitation of 100 mm/day over a 5-day period using daily and hourly time steps. Here, three kinds of absurdities arise in the daily step explicit Euler approximations. In the left column (sub-surface scaling parameter $m = 0{\cdot}025$ m), the store 'over-fills' on the first time step (i.e. a wet bias), and then spectacularly 'over-drains' in the second step—despite still receiving the same amount of rainfall as that in the first step. This causes a large pulse of baseflow and the storage does not recover within the simulation period despite all subsequent precipitation. In the middle column ($m = 0{\cdot}03$), the store still over-fills on the first step and still over-drains in the second step, but the excessive drainage is not as dramatic as when $m = 0{\cdot}025$. The result, however, is still worrisome: the sawtooth pattern is again evident, characterized by successive

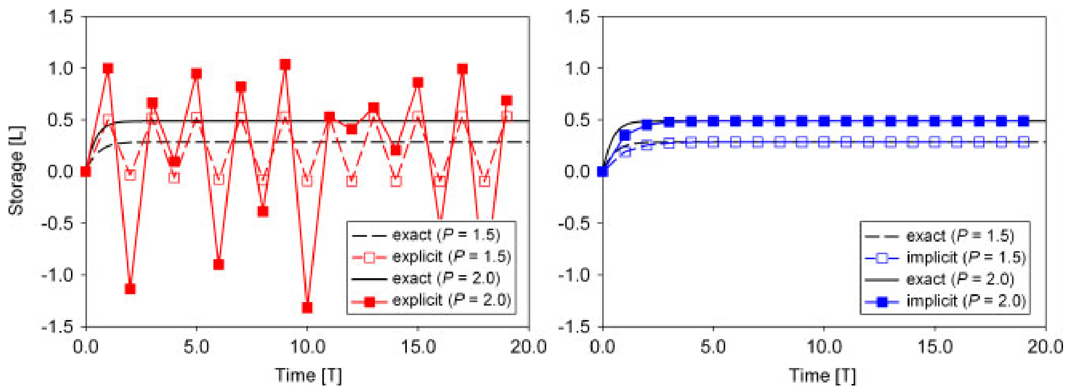

Figure 1. Absurdities in the behaviour of a simple reservoir approximated using the fixed-step explicit Euler scheme. Here, when $\Delta t = 1$ (T), increasing the precipitation P to 2·0 (L/T) (filled red squares) results in the explicit Euler approximation of the exponential store described by Equation (1) oscillating between wetter and drier values than when $P = 1{\cdot}5$ (L/T) (empty red squares). Yet, the exact solution (black lines) and the more robust [though still fixed-step with $\Delta t = 1$ (T)] implicit Euler approximation (blue squares) are smooth and well-behaved. The model parameters are fixed at $k = 1{\cdot}42$ (L^{-1}), $b = 1$ (L/T) and the initial condition $S_0 = 0$ (L)

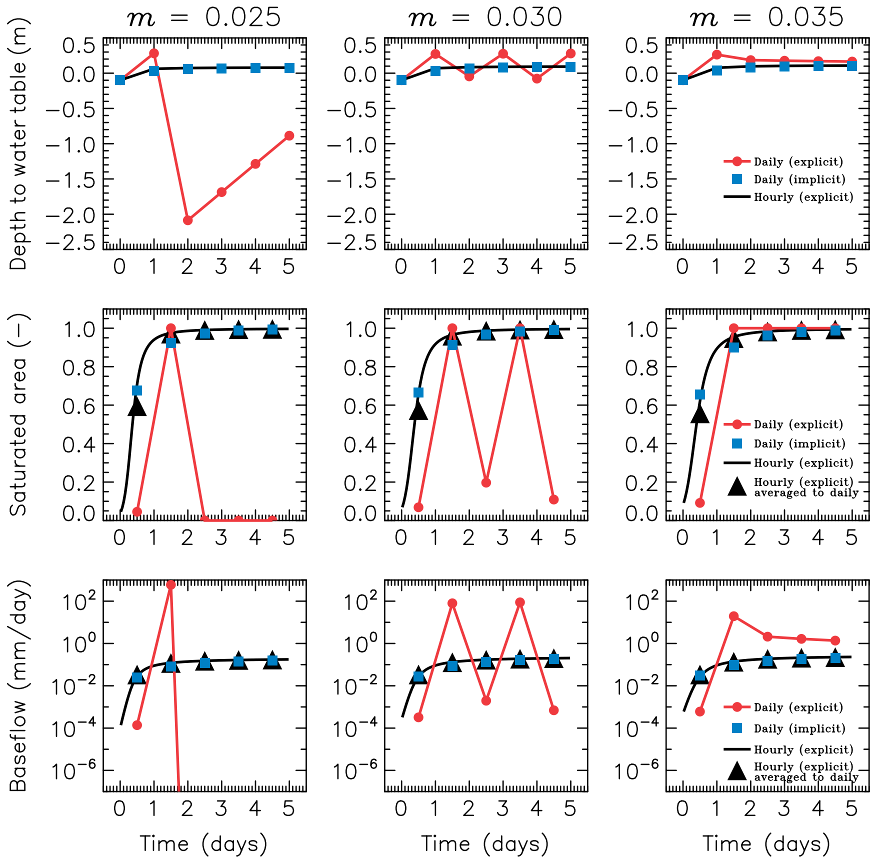

Figure 2. A more complex non-linear model behaving badly. Here, a TOPMODEL-like formulation is forced with constant precipitation (100 mm/day) and approximated using the explicit Euler scheme (daily and hourly time steps) *versus* the implicit Euler scheme (daily time steps). The figure shows simulations of the depth to the water table (top row), saturated area (middle row) and baseflow (bottom row), for different values of the sub-surface scaling parameter *m* (in metres). Note that the daily-step estimates of saturated area and baseflow are viewed as step-average values and hence the corresponding symbols are plotted midway through the step. For the hourly step results, the values averaged to the daily scale are shown for commensurability with the daily-step results. For a constant forcing, the model should smoothly approach a steady state (e.g. as seen in the hourly step solution). Yet, just like in simpler models, spurious sawtooth oscillations develop in the fixed-step explicit Euler approximation unless the step size is sufficiently small. Rainfall variability and/or additional model complexity could mask these oscillations, making them harder to detect, and allowing them to continue stealthily equivocating the unsuspecting hydrologist. And just like in the simpler example, possible cures include using the more robust implicit Euler approximation and/or using adaptive sub-stepping solutions

over-filling and over-draining. In the right column ($m = 0{\cdot}035$), the store also over-fills on the first time step as in the other two cases, but then drains more slowly.

In all the cases, the time series of storage, saturated areas and baseflow are markedly different in the daily *versus* the hourly runs. Also, while the daily step simulations are highly sensitive to the value of parameter m, the hourly step results are quite insensitive to m. Making physical sense of this behaviour is difficult—it could suggest a poorly conceptualized drainage flux, or intrinsic time scale dependencies, or both. Yet, the problem is neither: using the unconditionally stable implicit Euler approximation with fixed daily steps produces a solution with no oscillations. Its *only* difference from the explicit Euler scheme is that the solution is forced to satisfy the fluxes at the *end* of the time step (e.g., Press *et al.*, 1992, pages 728–730; see also Butcher, 2008).

More generally, in response to the opening question of this section, it is precisely the potential for compensatory interactions between the components of more complex models that *mandates* that each model component be numerically robust. Failing so, the process identifiability problems confronting the hydrologist will be further exacerbated by numerical approximation errors interacting with process representations and each other.

The full gamut

More recent work, combining numerical analysis with empirical evaluations of many different model structures across multiple catchments, suggests that numerical troubles in conceptual hydrology are more than just a couple of broken equations (Kavetski *et al.*, 2003; Clark and Kavetski, 2010). These studies have indicated that the numerical errors of uncontrolled time-stepping schemes routinely dwarf the structural

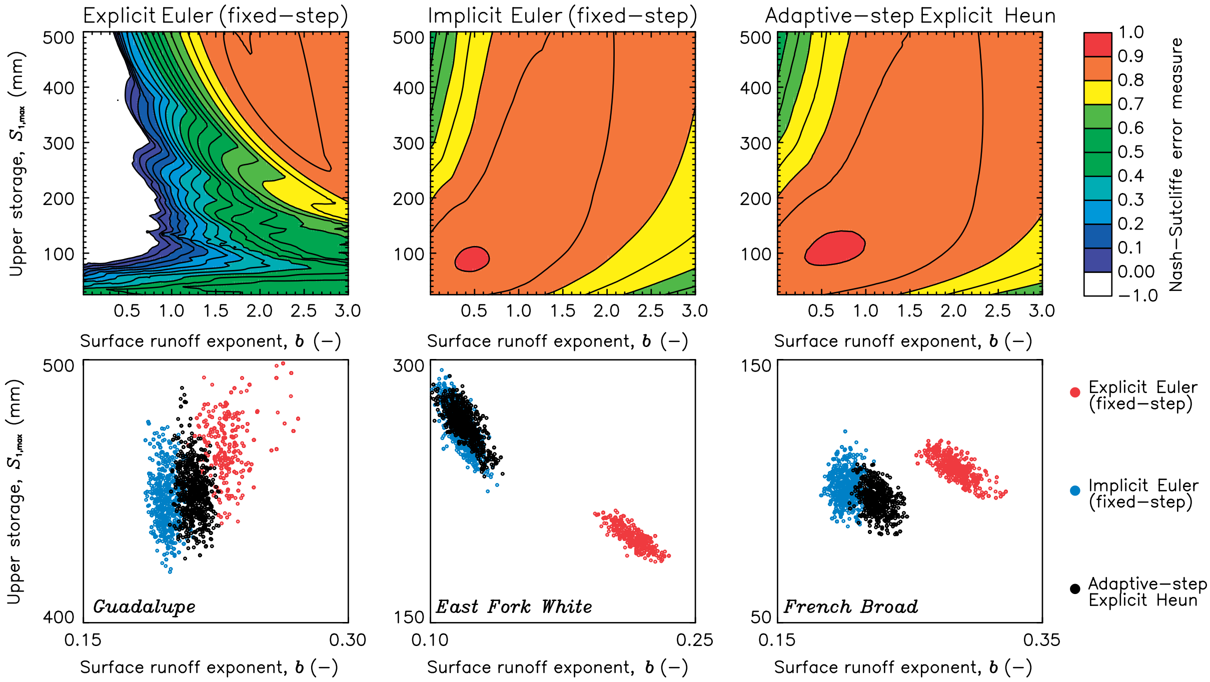

Figure 3. Selected numerical daemons of conceptual hydrological modelling: numerical errors in the model predictions cause micro- and macro-scale deformations of objective functions (top row) and lead to corrupted parameter inference (bottom row). The top row compares the Nash–Sutcliffe surface for a 2D slice through the model parameter space for explicit Euler, implicit Euler and an adaptive time stepping approximations of *the same model equations* applied in the Mahurangi River basin in New Zealand. The panels in the bottom row contrast MCMC-derived parameter estimates obtained for different numerical solutions of the same models applied in three different MOPEX basins (Duan *et al*., 2006). The differences between the MCMC-estimated parameter distributions within each panel in the bottom row arise *solely* from differences in the time-stepping scheme. Although in some cases numerical effects may appear to be minor (Guadalupe), this cannot justify complacency: in the absence of quality control (adaptive time stepping) or at least unconditional stability (implicit Euler), there is simply no guarantee that a different forcing regime and/or different catchment conditions (e.g. as seen in the East Fork White basin) will not bring out the worst in unreliable numerical methods such as the fixed-step explicit Euler scheme. Figure reproduced using the model output presented by Kavetski and Clark (2010), where further monstrosities are showcased

errors of the model conceptualization. This has serious implications for model analysis and predictive use, including inconsistent inferences of parameters and internal states even if the calibrated streamflow predictions are similar. Some of these effects can be staggering, as shown in Figure 3. Even when numerical errors fortuitously compensate for data and structural errors during calibration, they make the model unduly fragile in predictive mode, as evidenced in validation tests. Sensitivity analyses are also corrupted, in some cases, to the extent that they reflect the sensitivity of numerical errors rather than of the hydrological model itself (Kavetski and Clark, 2010). A disconcerting consequence is that the interpretation of hydrological model output to gain insights into internal catchment dynamics, including the relative significance and behaviour of different processes, can be severely compromised.

The lack of attention to numerical reliability may have also hindered progress in the prediction in ungauged basins (PUB) initiative [which, as noted by Sivapalan *et al*. (2003), is of high practical relevance as most basins around the world are ungauged]. For example, appropriate model structures and parameter values for ungauged catchments can, in principle, be derived from the analysis and interpretation of models calibrated in well-instrumented watersheds. However, as listed above, the conclusions of such studies can easily be tainted by numerical artefacts and hence may not provide a sound basis for interpretation and regionalization.

Given the endemic nature of numerical problems, the hydrological community should take note: its numerical daemons can be unforgiving! Even MCMC methods, touted for their strength in exploring complex multimodal distributions, are not immune, with a demonstrable degradation of both accuracy and efficiency (Kavetski and Clark, 2010; Schoups *et al*., 2010). Indeed, Schoups *et al*. (2010), in Figure 3 and related discussion, indicate that even sophisticated MCMC methods such as DREAM are slow to converge, or fail to converge, when applied to the geometrically complex objective functions of poorly implemented models [as illustrated in Figure 3 here, see also Figure 1 in Kavetski and Clark (2010); and Figure 1 in Kavetski *et al*. (2003)]. And how can MCMC schemes be immune? Even the most sophisticated MCMC algorithm cannot detect when the hydrological model itself is solved inaccurately. Moreover, we stress that even a (practically unattainable) *perfect*

MCMC sampler, if applied to a poorly implemented hydrological model, would merely reproduce parameter distributions deformed by numerical errors. In this respect, improving MCMC sampling algorithms does not in itself address the root cause of the problem—and indeed could simply mask it.

More Robust Approaches for Model Building and Analysis

A broader perspective on previous work

The earlier studies by Michel *et al.* (2003, 2005) attributed the troublesome model behaviour to 'mistakes' in the model-building process [e.g. 'mistaking' an instantaneous flux for a flow volume over a finite time step, as suggested by Michel *et al.* (2003), p. 113]. However, the straightforward use of fluxes to obtain flow volumes is consistent with numerical approximations, which necessarily assume constant, or otherwise, analytically simple (e.g. linear, quadratic, etc.) fluxes over finite time steps. Hence, this commentary argues for a subtly different, but arguably more general, perspective: these troublesome problems represent uncontrolled numerical approximation errors.

When viewed from this more general perspective, problems such as those reported by Michel *et al.* (2003, 2005) are not so much 'mistakes' in integrating the exponential store in TOPMODEL or the quadratic SCS-CN store, but rather are indicative of a broader problem of poor numerical implementation in conceptual hydrology. Indeed, numerical problems are not limited to exponential or quadratic stores: analogous pathologies, e.g. 'the initially wetter store becoming drier' and 'a store becoming drier as a result of more rain' readily arise in uncontrolled approximations of other non-linear reservoirs (Kavetski, 2005). Because of the dependence of the outflow on the unknown actual storage, flow volume and storage estimates based on a finite number of flux evaluations may not adequately reflect variations throughout the time step. Although such errors arise in any numerical approximation (including implicit schemes), fixed-step explicit schemes are known to be especially fragile. Indeed, while non-linearities in flux-state relationships can drastically exacerbate numerical artefacts, even linear reservoirs, when approximated without due regard to numerical errors, can produce meaningless unstable results. Hence, the hydrological community should tackle the issue of numerical errors in a general hydrological modelling context, rather than trying to 'mend' specific cases such as exponential and quadratic stores, etc.

So, what to do? Use more robust numerics!

When reading many classic hydrological papers [e.g. the derivation of the Probability Distribution Model (PDM) model by Moore and Clarke (1981)], one is often struck by their mathematical elegance: the problem is clearly formulated and analytical solutions used. However, nature is not always analytically tractable. Even combinations of simple individual fluxes are generally analytically non-integrable: e.g. $dS/dt = P$ and $dS/dt = -kS^{\alpha}$ both have analytical solutions, yet $dS/dt = P - kS^{\alpha}$ does not. Restricting hydrology to analytical models would prevent adequate representations of many important non-linear processes (Dunne and Black, 1970; Western *et al.*, 2004; Tromp-van Meerveld and McDonnell, 2006).

Yet, reliable and well-established techniques are available for the approximate solution of differential equations, especially for the comparatively simple cases arising in conceptual hydrological modelling [e.g., see Butcher (2008) for good review of the extensive field of numerical integration]. For example, the implicit Euler time-stepping scheme avoids unstable growth of errors and is widely used in groundwater, petroleum, geotechnical and other engineering software (Clark and Kavetski, 2010). Even more reliable are adaptive time stepping schemes, which reduce the step size in regions of rapid change of the solution (e.g. in response to a strong forcing such as a rainfall event) but lengthen it during quiescent periods. More than just preventing instabilities, adaptive methods can constrain numerical approximation errors below a user-prescribed tolerance and are standard in mathematical software, including both commercial packages, public libraries and multi-physics packages [e.g., Shampine and Reichelt (1997) and many others].

When applied to conceptual hydrological models, the use of robust time-stepping schemes proved immediately beneficial. Unconditionally stable implicit schemes, as well as adaptive methods with error control, not only significantly simplify the model calibration problem in terms of accuracy and efficiency but also markedly broaden the repertoire of model analysis techniques, including hitherto abandoned Newton-type optimization (Kavetski and Clark, 2010). We do note that some subtleties arise when selecting time-stepping schemes in the context of derivative-based parameter optimization, sensitivity analysis and uncertainty estimation. For example, fixed-step implicit schemes, despite lacking strict error control, may be favoured over adaptive integration because they more readily support numerically smooth model response surfaces (Gill *et al.*, 1981). We also stress that while robust time stepping will remove spurious artefacts, genuine model non-linearities may still give rise to potentially complicated response surfaces and objective functions [Figure 1 in Kavetski *et al.* (2006)].

Computational 'inefficiencies' and programming complexities

It is sometimes claimed that accurate time stepping, in particular employing adaptive and/or implicit methods, is 'computationally inefficient'. This confuses 'efficiency' and 'cost'. A numerically efficient algorithm achieves high accuracy at minimal cost. In this respect, a fixed-step explicit Euler solution is not 'computationally efficient'—it is cheap and inaccurate! In our opinion, while hydrologists, like any scientists, will always strive to tackle problems at the edge of and beyond computational feasibility, it is imprudent to exchange reliability for speed, even if it requires more attention to code development and, in some cases, the use of more specialized techniques and packages.

It is worth noting that robust numerical solutions—including algorithm selection and implementation—are not *always* straightforward, especially for complex spatially distributed models. Moreover, while implementation costs are relatively low for well-designed code exploiting the state-space form of the model equations, retrofitting an existing ad hoc code may require substantial code re-organization and enhancement. Nevertheless, we stress that cases where method selection is critical and non-trivial are often precisely those for which simplistic uncontrolled techniques are most unreliable, and for which a robust carefully designed algorithm is sorely needed!

Better Hypothesis Testing: Separation of Concepts from Implementation

It could be argued that much of the original motivation for numerical methods came from engineering. While mathematicians tended to look for an 'exact solution to an approximate problem', engineers often preferred 'approximate solutions to the exact problem'. This can create considerable confusion in the context of hypothesis testing—because when scientists talk about 'hypothesis testing', they usually refer to the representation of some phenomenon, rather than to the accuracy with which some equation is solved.

Hence, a clear delineation of the governing equations from their numerical implementation provides better support for hypothesis testing. It allows hydrologists to focus their attention on improving model representations of nature, without confusing process conceptualization ('structural') errors with numerical errors of the selected solution technique. Importantly, multiple alternative representations of nature can and should be evaluated within the same robust numerical framework (Clark *et al*., 2008). It is stressed that while numerical accuracy (even analytical solutions) cannot, even in principle, reduce the structural errors in the hypothesized governing equations, numerical robustness plays the irreplaceable role of eliminating *unnecessary* spurious artefacts. It also avoids compensation of numerical errors by model parameter values (see Figure 11 in Kavetski and Clark, 2010), which is important if the hydrologist is interested in 'getting the right answers for the right reasons' (Kirchner, 2006).

In our opinion, the numerical solution should not be viewed as part of the model itself. For example, interpreting the operator-splitting approximation (tacitly used in common hydrological models such as GR4J, VIC and Sacramento model, see Kavetski *et al*., 2003; Schoups *et al*., 2010) as a genuine model feature is conceptually unsatisfying: it implies that fluxes operate sequentially *within a single time step*—e.g. surface runoff, followed by baseflow and evaporation. Yet we know that these processes take place concurrently. These considerations are relevant in numerical design: unless adaptive sub-stepping is incorporated into the operator-splitting algorithm (Schoups *et al*., 2010), the dependence of the solution on the selected sequence of fluxes makes the computer implementation reliant on rather arbitrary assumptions.

Closing Remarks

A numerically savvy reader may wonder 'what is all the fuss about?' Is it not obvious that differential equations must at least be solved stably, and that reliable error control is part of most ordinary differential equation (ODE) solvers? Has this point not been already made countless times in the applied ODE literature (Butcher, 2008), in the 'popularly oriented' numerical texts (Kahaner *et al*., 1989; Press *et al*., 1992) and even in software manuals [e.g. the Matlab ODE toolkit by Shampine and Reichelt (1997) simply disallows fixed-step integration]? We do indeed get these legitimate questions from colleagues in more numerically oriented disciplines, such as groundwater, physics, mathematics and others.

We also note that many hydrological models have already been implemented using robust techniques. These include, for example, up-scaling (Reggiani and Rientjes, 2005), regionalization (Kling and Gupta, 2009), residence time analysis (Vache and McDonnell, 2006) and other applications. In conceptual rainfall–runoff modelling, unified state-space formulations and robust numerics are exploited in the Framework for Understanding Structural Errors (FUSE) (Clark *et al*., 2008). In the broader environmental literature, we can point to groundwater (Harbaugh, 2005), geochemical (Binning and Celia, 2008) and other applications. In some studies, 'heuristic' time stepping has been used, for example, selecting 'normal', 'medium' and 'small' time steps depending on the thickness of snow layers (Marks *et al*., 1999), varying the time step in a rainfall–runoff model depending on the observed rainfall forcing (Hughes and Sami, 1994), setting the time step size based on the estimated time of concentration of a catchment (Maniak, 1997), etc. Though such heuristics may lack the stringency and reliability of formal error control [e.g. see Figure 4 in the Richards equation case study by Kavetski *et al.* (2002)], they may at least guard against gross numerical errors and absurdities.

Yet we argue that the robust numerics paradigm is yet to be accepted as a *required standard* in conceptual hydrology. For example, the collections listed in Singh and Frevert (2002a,b), and most other studies, fail to report the techniques used to solve their model equations, making their numerical robustness difficult to ascertain. This can be asking for trouble: confounding numerical artefacts are not some rare isolated instances, but can affect virtually any model structure, in any catchment, and under common hydroclimatic conditions (Clark and Kavetski, 2010; Schoups *et al.*, 2010; see also earlier work by Kavetski *et al.*, 2003; Michel *et al.*, 2003, and others). These are the key concerns raised in this commentary.

We also note that, in the same breath as urging new ways of 'holistic' thinking and training, it is sometimes suggested that the hydrological community shift away from 'traditionalist' technical training (which, in engineering, should include linear algebra, differential equations, and so forth). We see little direct advocacy for the theme of robust numerics in leading commentaries on the future of hydrology (e.g., Wagener *et al.*, 2010). Nor is the often-poor attention to robust mathematics recognized as a major factor 'undermining the science of hydrology' by other prominent commentaries (e.g., Andreassian *et al.*, 2007; Beven, 2008; Buytaert *et al.*, 2008; Sivapalan, 2009; and others). It is indisputable that much innovation is needed in the pursuit of hydrological process understanding and its incorporation into models. Yet, in our quest towards new imaginative hydrological solutions, we must not overlook the fact that they will, one way or another, be numerically based. Then, a cavalier attitude to the underlying mathematics—whether due to tacit disregard or lack of training—can easily create veritable numerical daemons that will thwart scientific and operational progress.

Hence, while certainly not suggesting that the hydrologist and experimentalist must become numericians, this commentary stresses that they must nonetheless be comfortable with formulating their insights in clear mathematical terms (e.g. in state-space form) and with employing robust, quality-controlled numerical techniques in model development and analysis—whether by programming themselves, investing into canned toolkits, or, perhaps as a last resort, by inviting the Numerician into the hitherto tête-à-tête dialogue between the Modeller and the Experimentalist (Seibert and McDonnell, 2002).

Reliable hydrological modelling is a major challenge without adding confounding numerical artefacts—there remain the formidable unresolved issues of data uncertainty and structural error treatment. We stress that, unlike other thorny issues for which a consensus is far from established (see discussions in Beven *et al.*, 2008; Renard *et al.*, 2010; Doherty and Welter, 2010; and others), numerical approximations are well-understood and uncontroversial for the types of equations usually encountered in conceptual hydrology. Continuing to rely on the simple explicit scheme, the archetypal numerical integration method introduced 240 years ago by the prodigious Leonhard Euler (1768–1770), fails to exploit major mathematical developments of the last two centuries and is hardly defensible. In the absence of robust numerical technology, seemingly trivial numerical aspects can have profound impact on the model performance, equivocating scientific analyses and undermining operational predictions. We hope that this commentary will further motivate the Hydrologist to avoid preventable numerical troubles before tackling more genuine challenges.

Acknowledgements

We thank the Editor Jim Buttle for the opportunity to present this commentary. The constructive discussions with Keith Beven, Philip Binning, Michael Celia, James Craig, Fabrizio Fenicia, Stewart W. Franks, Hoshin Gupta, Ethan Guttmann, Tom Hopson, Bethanna Jackson, George Kuczera, Mark Masia, Jan Nordbotten, Manfred Ostrowski, Mark Thyer, Andrew Slater, Scott W. Sloan, Thorsten Wagener, Jasper Vrugt and Eric Wood on earlier incarnations of this work are gratefully acknowledged.

References

Andreassian V, Lerat J, Loumagne C, Mathevet T, Michel C, Oudin L, Perrin C. 2007. What is really undermining hydrologic science today? Hydrological Processes 21: 2819–2822.

Beven K. 1997. TOPMODEL: a critique. *Hydrological Processes* 11(9): 1069–1085.

Beven K. 2008. On doing better hydrological science. *Hydrological Processes* 22: 3549–3553.

Beven KJ, Binley AM. 1992. The future of distributed hydrological models: model calibration and uncertainty prediction. *Hydrological Processes* 6: 279–298.

Beven KJ, Smith PJ, Freer JE. 2008. So just why would a modeller choose to be incoherent? *Journal of Hydrology* 354(1–4): 15–32.

Binning PJ, Celia MA. 2008. Pseudokinetics arising from the upscaling of geochemical equilibrium. *Water Resources Research* 44: W07410. DOI: 07410.01029/02007WR006147.

Butcher J. 2008. *Numerical Methods for Ordinary Differential Equations*, 2nd edn. Wiley: Chichester.

Buytaert W, Reusser D, Krause S, Renaud J-P. 2008. Why can't we do better than Topmodel? *Hydrological Processes* 22: 4175–4179.

Cherkauer KA, Bowling LC, Lettenmaier DP. 2003. Variable infiltration capacity cold land process model updates. *Global and Planetary Change* 38(1–2): 151–159.

Clark MP, Kavetski D. 2010. Ancient numerical daemons of conceptual hydrological modeling. Part 1: Fidelity and efficiency of time stepping schemes. *Water Resources Research* 46: W10510, DOI: 10.1029/2009WR008894.

Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, Wagener T, Hay LE. 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research* 44: W00B02. DOI: 10.1029/2007WR006735.

Doherty J, Welter D. 2010. A short exploration of structural noise. *Water Resources Research* 46: W05525. DOI: 05510.01029/02009WR008377.

Duan Q, Schaake J, Andreassian V, Franks SW, Goteti G, Gupta HV, Gusev YM, Habets F, Hall A, Hay L, Hogue T, Huang M, Leavesley G, Liang X, Nasonova ON, Noilhan J, Oudin L, Sorooshian S, Wagener T, Wood EF. 2006. Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. *Journal of Hydrology* 320(1–2): 3–17.

Duan QY, Sorooshian S, Gupta V. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* 28(4): 1015–1031.

Dunne T, Black RD. 1970. Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research* 6: 1296–1311.

Euler L. 1768–1770. *Institutiones Calculi Integrals Volumen Primum (MCMXIII)*. http://www.archive.org/details/institutionescal020326mbp, [11 September 2010], published by Lipsiae et Berolini.

Gill PE, Murray W, Wright MH. 1981. *Practical Optimization*. Academic Press: London.

Harbaugh AW. 2005. MODFLOW-2005, the U.S. Geological Survey modular ground-water model—the Ground-Water Flow Process, U.S. Geological Survey Techniques and Methods 6-A16.

Hendrickson JD, Sorooshian S, Brazil LE. 1988. Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. *Water Resources Research* 24(5): 691–700.

Hughes DA, Sami K. 1994. A semi-distributed, variable time interval model of catchment hydrology—structure and parameter estimation procedures. *Journal of Hydrology* 155: 265–291.

Ivanov VY, Vivoni ER, Bras RL, Entekhabi D. 2004. Catchment hydrologic response with a fully distributed triangulated irregular network model. *Water Resources Research* 40(11): W11102. DOI: 11110.11029/12004WR003218.

Kahaner D, Moler C, Nash S. 1989. *Numerical Methods and Software*. Prentice-Hall: Englewood Cliffs, NJ.

Kavetski D. 2005. Analysis of input data uncertainty and numerical robustness in conceptual rainfall-runoff modelling. PhD thesis, Faculty of Engineering, University of Newcastle, Australia.

Kavetski D, Binning P, Sloan SW. 2002. Adaptive backward Euler time stepping with truncation error control for numerical modelling of unsaturated fluid flow. *International Journal for Numerical Methods in Engineering* 53(6): 1301–1322.

Kavetski D, Clark MP. 2010. Ancient numerical daemons of conceptual hydrological modeling. Part 2: Impact of time stepping schemes on model analysis and prediction. *Water Resources Research* 46: W10511, DOI: 10.1029/2009WR008896.

Kavetski D, Kuczera G, Franks SW. 2003. Semidistributed hydrological modeling: a "saturation path" perspective on TOPMODEL and VIC. *Water Resources Research* 39(9): 1246. DOI: 1210.1029/2003WR002122.

Kavetski D, Kuczera G, Franks SW. 2006. Calibration of conceptual hydrological models revisited: 2. improving optimisation and analysis. *Journal of Hydrology* 320(1–2): 187–201.

Kirchner JW. 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* 42(3): W03S04. DOI: 10.1029/2005WR004362.

Kling H, Gupta H. 2009. On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability. *Journal of Hydrology* 373: 337–351.

Maniak U. 1997. *Hydrologie und Wasserwirtschaft: Eine Einführung für Ingenieure*, 4th edn. Springer: Heidelberg, Germany.

Marks D, Domingo J, Susong D, Link T, Garen D. 1999. A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes* 13: 1935–1959.

Michel C, Andreassian V, Perrin C. 2005. Soil conservation service curve number method: how to mend a wrong soil moisture accounting procedure. *Water Resources Research* 41: W02011. DOI: 02010.01029/02004WR003191.

Michel C, Perrin C, Andreassian V. 2003. The exponential store: a correct formulation for rainfall-runoff modelling. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* 48(1): 109–124.

Moore RJ, Clarke RT. 1981. A distribution function approach to rainfall runoff modeling. *Water Resources Research* 17(5): 1367–1382.

Perrin C, Michel C, Andreassian V. 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* 279(1–4): 275–289.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1992. *Numerical Recipes in Fortran-77: The Art of Scientific Computing*. University of Cambridge: Cambridge; 820.

Reggiani P, Rientjes THM. 2005. Flux parameterization in the representative elementary watershed approach: Application to a natural basin. *Water Resources Research* 41: W04013. DOI: 04010.01029/02004WR003693.

Renard B, Kavetski D, Thyer M, Kuczera G, Franks SW. 2010. Understanding predictive uncertainty in hydrologic modeling: Le challenge of identifying input and structural errors. *Water Resources Research* 46: W05521. DOI: 05510.01029/02009WR008328.

Schoups G, Vrugt JA, Fenicia F, van de Giesen NC. 2010. Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models. *Water Resources Research* 46: W10530, DOI: 10.1029/2009WR008933.

Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research* 38(11): 1241. DOI: 1210.1029/2001WR000978.

Shampine LF, Reichelt MW. 1997. The MATLAB ODE suite. *SIAM Journal on Scientific Computing* 18(1): 1–22.

Singh VP, Frevert DK. 2002a. *Mathematical Models of Large Watershed Hydrology*. Water Resources Publications: Highlands Ranch, Colorado.

Singh VP, Frevert DK. 2002b. *Mathematical Models of Small Watershed Hydrology and Applications*. Water Resources Publications: Highlands Ranch, Colorado.

Sivapalan M. 2009. The secret to "doing better hydrological science": change the question!. *Hydrological Processes* 23: 1391–1396.

Sivapalan M, Beven K, Wood EF. 1987. On hydrologic similarity: 2. a scaled model of storm runoff production. *Water Resources Research* 23: 2266–2278.

Sivapalan M, Takeuchi K, Franks SW, Gupta VK, Karambiri H, Lakshmi V, Liang X, McDonnell JJ, Mendiondo EM, O'Connell PE, Oki T, Pomeroy JW, Schertzer D, Uhlenbrook S, Zehe E. 2003.

IAHS decade on Predictions in Ungauged Basins (PUB). *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* 48(6): 857–880.

Tromp-van Meerveld HJ, McDonnell JJ. 2006. Threshold relations in subsurface stormflow: 2. the fill and spill hypothesis. *Water Resources Research* 42: W02411. DOI: 02410.01029/02004WR003800.

Vache KB, McDonnell JJ. 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. *Water Resources Research* 42: W02409. DOI: 02410.01029/02005WR004247.

Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39(8): 1201. DOI: 1210.1029/2002WR001642.

Vrugt JA, ter Braak CJF, Diks CGH, Robinson BA, Hyman JM, Higdon D. 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation* 10(3): 273–290.

Wagener T, Sivapalan M, Troch PA, McGlynn BL, Harman CJ, Gupta HV, Kumar P, Rao PSC, Basu NB, Wilson JS. 2010. The future of hydrology: An evolving science for a changing world. *Water Resources Research* 46: W05301. DOI: 05310.01029/02009WR008906.

Western AW, Zhou SL, Grayson RB, McMahon TA, Bloschl G, Wilson DJ. 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology* 286: 113–134.