HYDROLOGICAL PROCESSES
Hydrol. Process. **21**, 2447–2457 (2007)
DOI: 10.1002/hyp.6410

Simulating watershed runoff with a new data model

Femke Reitsma[1]* and Ralph Dubayah[2]

[1] *Institute of Geography, School of Geosciences, The University of Edinburgh, Drummond Street, Edinburgh, EH8 9XP, UK*
[2] *Department of Geography, University of Maryland, College Park, MD 20742, USA*

Abstract:

This paper presents the implementation of a new data model for simulating watershed runoff. This data model, called *nen*, allows users to visualize and analyse the processes that are simulated in the watershed runoff model. It extends data models typically used in modelling watershed runoff, such as raster, that do not give direct insight into the spatial dynamics and distribution of the processes expressed in the model. Data from the Upper Sheep Creek sub-watershed in the Reynolds Creek Experimental Watershed was used to illustrate the methodology within the context of runoff simulation. Model results, while consistent with standard approaches to validation against a hydrograph, require new methods of validation for testing the process descriptions and simulated process results that are unique outputs of this model. Future developments aim at creating analytical and validation techniques and further developing the simulator to allow for larger scale modelling. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS process; modelling; representation; watershed runoff; conceptual model; data model; GI Science

Received 23 June 2005; Accepted 7 March 2006

INTRODUCTION

There are numerous approaches to modelling watershed runoff, which range from simple to complex dynamic models with many parameters. Regardless of the level of detail captured in the dynamic model, only a limited set of data models are available for representing the results. Within a simulation environment, data models are used to represent the dynamics of the environmental model and store its results. These data models have a direct impact on what can and cannot be explained from the results of a simulation.[1]

The difficulty with the data models typically used in modelling watershed runoff, such as raster, is that they do not give direct insight into the spatial dynamics and distribution of the processes expressed in the dynamic model. For example, understanding the distribution of Hortonian overland flow with current data models requires an in depth knowledge of how the dynamic model operates and the relationship between the results at each time step and the processes that are included in the model. However, there is not a one to one relationship between the modelled pattern of results and the processes that caused them, as expressed by the concept of equifinality. There are multiple process pathways to the same result. Current data models present the future state of the system but do not express how that future state of the system came about, that is, they do not express the processes that caused those states. With an appropriate data model, we can capture and analyse information about the processes that are formalized in the dynamic model.

The objective of this paper is to describe a watershed runoff model and present its results within the context of a new data model that represents processes. The advantages of the methodology described in this paper are that it allows for the querying, analysis, and exploration of the processes represented in the dynamic model. This not only provides insight into how well our dynamic model functions by allowing for the visual representation of process dynamics, but furthers our understanding and explanation of the processes modelled. This is in contrast to current approaches that utilize data models to represent the state of the system at an instant of time, which gives no information about the modelled processes. Hence, the underlying argument of this paper is that the simulation method applied to a dynamic model has epistemological implications.

The remainder of this paper is organized as follows. First, existing data models used in watershed modelling are reviewed. Next, the study area and data is described, namely Reynolds Creek Experimental Watershed, which is followed by a discussion of an alternative data model that is used to simulate a dynamic model of watershed runoff for that data set. The methodology is then discussed, including details of the dynamic model and simulator that were developed to implement the new data model. We next present the model results, showing a conventional comparison with a stream hydrograph, as well as the unique outputs of our simulation that allows the spatio-temporal characteristics of the associated runoff

---

* Correspondence to: Femke Reitsma, Institute of Geography, School of Geosciences, The University of Edinburgh, Drummond Street, Edinburgh, EH8 9XP, UK. E-mail: femke.reitsma@ed.ac.uk

[1] Here the term model is being used to describe the formal specification or representation of the system being modelled, and simulator as the environment in which the model is run. In order to clarify the distinction between model and data model, the term dynamic model is used throughout.

processes to be examined. Finally, we discuss the difficulties and potential advantages of implementing and validating a dynamic model with the nen data model for larger scale application.

DATA MODELS IN WATERSHED SIMULATIONS

There are many ways of classifying dynamic models both in general and within the field of hydrology. For example, dynamic models may be classified by conceptual type, such as empirically or physically based models, or by spatial type, such as lumped or spatially distributed models (Singh, 1995; Grayson and Blöschl, 2000; Beven, 2001; Mulligan, 2004). The slightly different approach taken here is to consider them according to data modelling primitives, in order to clearly state the case for the value of an alternative data model for representing processes before presenting the watershed model and its results. In particular, distributed models are the focus as they explicitly represent the spatial nature of hydrological processes.

Models traditionally come in two basic forms, distributed and spatially distributed models. Distributed modelling divides the watershed into discrete spatial units, computes the response of each unit to inputs such as precipitation, and then combines them to give the response for the entire watershed, such as the SHE model (Abbott *et al.*, 1986a,b). This approach does not capture the spatial interaction of processes at or between spatial or temporal scales, and '[i]n some respects the distributive mechanism means that the distributed model is essentially a 'lumped' model at grid scale' (Ward and Robinson, 2000: 348). In contrast, spatially distributed modelling explicitly deals with interactions among neighbouring spatial units. Such dynamic models are typically used to route water flow over a landscape using flow direction algorithms the simplest of which results in sending water to a neighbouring downslope element that has the greatest elevation decrease within its eight cell or Moore neighborhood, known as the D8 method (Endreny and Wood, 2003).

Most of these advanced, spatially distributed, rainfall-runoff models are based on the classic, physically based, distributed model blueprint designed by Freeze and Harlan (1969). This design describes a basic framework for numerical modelling, with a set of partial differential equations operating over a set of points arranged in a three-dimensional grid representing the watershed. The more recent framework provided by Beven does not break the data modelling mould of the original, rather it considers how different dynamic models and their parameters might fit into a model space (Beven, 2002). This data modelling approach is ubiquitous in spatially distributed models, that is, the use of the pixel, point, line, or polygon primitives, which at each instant of time in the dynamic model are described by a set of attributes. Pixels are used in grid representations of the watershed, specifying a value such as elevation at a specific location; points can be similarly used to represent a continuous field of data, or may be used to represent specific data collection points such as lysimeters or piezometers; lines are typically used to model flow networks; and polygons are used for representing larger areas of interest such as hydrological response units. The underlying general data structure for all of these primitives is defined by a spatial location x, y, z, a time point t, and a set of attributes $\{a_1, a_2 \ldots a_n\}$. Regardless of whether the equation is physically based, empirically based, or stochastic in nature, the underlying representational devices, the data models, remain the same.

Consequently, the types of output available to the dynamic model user, which lead to analysis and querying techniques, are also fundamentally the same. Although substantive output may vary from one dynamic model to another dynamic model, such as whether or not sediment or chemistry is modelled (Borah and Bera, 2004, 2003), the structure of the information provided is consistently the state of that output at each instant of time. For example, TOPMODEL is a spatially distributed model that uses an index of similarity called the topographic index to define its spatial units, and uses a flow routing algorithm to direct water through these units (Beven, 2001). The output of TOPMODEL predicts watershed discharge and the spatial distribution of saturation at any temporal instant in the simulation or as a cumulative output at the end of the simulation.

Since the mid-1990s, there has been an increasing amount of research and development on the integration of hydrological models and Geographical Information Science (GIS) (Romanowicz and Beven, 1993; Streit and Wiesmann, 1996; Feng and Sorokine, 2001). This linkage of GIS and hydrological modelling has ranged from loose coupling, which is simply the transfer of data from one program to the other, to hydrological models embedded within a GIS, such as the LISFLOOD model developed within PCRaster by De Roo *et al.* (2000). This integration has aided dynamic modelling by easing problems of spatial data input and by tapping into the data management and analytical tools of GIS. Yet, as with earlier models, the underlying data models remain the same, which as expressed by Maidment are three basic (point, line, polygon) and three derived (grid, triangular irregular network (TIN), network) data models (Maidment, 1993).

Cellular Automata (CA) presents an alternative approach for modelling spatially continuous phenomena. Recent advances in modelling geomorphological change use CA, which extend the spatially distributed modelling approach (e.g. Coulthard *et al.*, 2000; Favis-Mortlock *et al.*, 2000; Haff, 2001; Pullar, 2003). Not only do the grid cells interact, for example, excess energy in one cell may be transported to a neighbouring cell based on a range of cell characteristics, but CA allow for the interaction between the structure of the landscape and the processes operating over it. However as with the earlier dynamic models, the data model used is still state based. Each cell in an application of CA contains information about the state of that cell at an instant of time. Thus the resultant dynamics of the model can only be interpolated between time slices.

The results of watershed models can be classified as temporal, such as the hydrograph, spatial, such as the accumulated spatial distribution of runoff, or spatio-temporal, such as the change in the spatial distribution of runoff over time. These results are used to compare and validate dynamic models (Veith *et al.*, 2003). With the currently available data models two types of query and analysis can be undertaken, firstly of the state of entities or parameters at an instant of time, or of the difference between states of the entity or parameter values at different time instances (for example van Oosterom *et al.*, 2002). For an example of the latter, Gao *et al.* (1993) use sequences of frames to show change in the distribution of attributes as physical fields.

AN ALTERNATIVE DATA MODEL

Interpolating processes

Representing dynamic model results with state-based data models, such as rasters or polygons, means that the process information represented in the dynamic model is lost or at best can only be interpolated between states at particular instants of time. The problem with interpolation is that, as with spatial information, the wrong process may be interpolated. Take, for example, the case provided by Baird (2004) who finds two quite reasonable yet distinctly different explanations of a pattern found in the output of a hydrological model. Baird observes that the temporal pattern of high initial flow rates in the soil, followed by a steep decline after a precipitation event, can be explained by two different processes, one being the importance of macropores in a model utilizing a combination of Darcy's law and the Richards equations, and the other by the entrainment of air bubbles which over time coalesce and block the flow of water. Representing processes explicitly provides the opportunity to explore, which processes are dominant and whether our descriptions of those processes are correct; the proposed data model does just that. As expressed by Mulligan, 'there are still areas in which the complexity of hydrological processes is so great, or the information so little, that we do not understand the processes well enough to develop reliable models' (2004: 117). If the processes are not understood, how can they be modelled, visualized, and explained in a model? The proposed approach allows for testing hypotheses about descriptions of processes.

For example, the Automated Geospatial Watershed Assessment tool (AGWA) is a hydrological watershed modelling tool that integrates a GIS and the existing hydrological models of KINEROS2 and soil and water assessment tool (SWAT). In their description of the AGWA tool, Miller *et al.* (2002) give examples of change detection in water yield, that is, supporting the visualization of increase or decrease in the spatial distribution of water runoff over time; yet this does not provide any insight into the processes that cause these changes. A further example of the difficulty of relating process to form is provided by Gurtz *et al.* (2003), who implemented and analysed the results of two models for the same catchment, namely WaSiM-ETH, a physically based and grid modelled water balance model, and PREVAH, a conceptual model based on hydrological response units. Despite the fact that these two models differ by assigning water flow to different processes, both models simulate watershed discharge realistically in comparison with observations.

The nen data model

Focusing on processes does not imply that a mere switch has been made from object or structure to process, from material substance to laws of systems dynamics. Rather the process data model is defined here to encapsulate both material substance and rules or physical laws that specify the behaviour of the process. As the single modelling primitive, process encapsulates both matter and movement into one 'thing'.

This representative entity forms the data modelling primitive for processes, which can be expressed in tuple form as $(x1, y1, x2, y2, st, \{a1, a2, \ldots\}, \{r1, r2, \ldots\})$, or graphically as a (node, edge, node) triple as illustrated in Figure 1. Each (node, edge, node) triple will be henceforth referred to as a *nen*. Note that this is only a representation of a point process, which proved to be the best representation for the watershed runoff model, but it can also be extended to areal or linear feature.

The location of the process is specified by $x1$, $y1$, $x2$, $y2$, which expresses the spatial extent of the process; this can be extended to the z-dimension. The st represents the spatio-temporal granularity of the process, which expresses how far and over what time period the process will operate, which may be a function of the amount of energy that initiates the process. The set $\{a1, a2, \ldots\}$ defines the set of attributes of the process. The set $\{r1, r2, \ldots\}$ defines the set of rules of the process that govern its dynamics and interaction with other processes and external inputs. For example, a set of rules for modelling the process of groundwater flow may define the spatio-temporal extent of an instance of that process as a number of mm/hour depending on various relationships it holds among other processes.

The advantage of using this data model in simulating a watershed model is that we can explore the spatial dynamics of the processes as well as more traditional outputs of the location of matter such as water or soil. The *nen* allows us to query for the location and

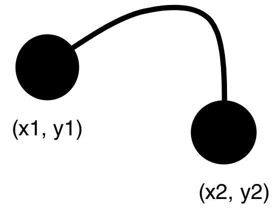

Figure 1. Process oriented data model, termed *nen*

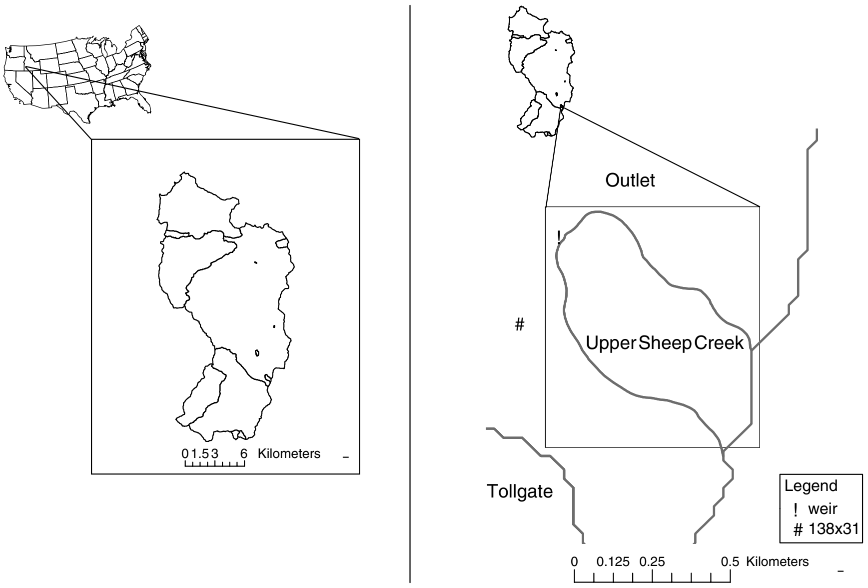

Figure 2. Location of the study area, Reynolds Creek Watershed (left), and Upper Sheep Creek watershed within Reynolds Creek Watershed (right)

attributes of processes at an instant of time, answering such queries as 'Where do certain processes dominate in our system?'. Furthermore, it allows us to gain a better understanding of how our dynamic model represents these processes and whether or not it does so realistically. This approach to modelling extends beyond modelling for purely predictive purposes to developing dynamic models where the goal is to accurately represent processes, thereby allowing the model to be used as an explanatory device in furthering knowledge about those modelled processes.

STUDY AREA AND DATA

The dataset used to develop the watershed modelling test case is from the Reynolds Creek Experimental Watershed (RCEW). This watershed has been well studied and is the basis of a high-quality long-term dataset that was recently released to the research community; it is available via anonymous ftp: ftp.nwrc.ars.usda.gov, and is maintained by the U.S. Department of Agriculture Agricultural Research Service, Northwest Watershed Research Center in Boise, Idaho, United States (http://www.nwrc.ars.usda.gov).

Seyfried *et al.* (2001a); Slaughter *et al.* (2001) and Marks (2001) provide a detailed description of the RCEW, which we summarize here. The RCEW is 239 km^2, ranging in elevation from 1101 m to 2241 m above mean sea level. It is located in the Owyhee Mountains of south-western Idaho, United States (Figure 2). Reynolds Creek, the stream draining the watershed, is a third-order perennial stream that drains north to the Snake River. Approximately 77% of the watershed is under public (federal or state) ownership, with the remainder being privately owned and utilized for livestock grazing with some irrigated fields along the creek at lower elevations. Within the RCEW there is large variation in local climate, geology, soils, and vegetation.

The spatial extent of the RCEW dataset was subset due to computational limitations; it also proved an easier test bed for development and exploration of the watershed model. Upper Sheep Creek, a small sub-watershed, provided such a subset to test the modelling approach (Figure 2). The primary characteristics of Upper Sheep Creek is a drainage area of 25·9 ha (digital elevation model (DEM) calculated), an elevation range of 1839–2017 m, and an intermittent streamflow regime. It was selected because it is the only small sub-watershed that can be best approximated by a rectangle, necessary due to the limitations of the simulation software; it has an intermittent regime rather than ephemeral, therefore it should produce more runoff; and it was included in a study that contains summary statistics on evapotranspiration, which was used in the model (Hanson and Wight, 1995).

The temporal simulation interval was restricted by precipitation records, the availability of streamflow data, and selection of a precipitation event that could be clearly mapped to a discharge event. The precipitation data are continuous records available for 12 sites, 20–32 year records available for 8 sites, 10–19 year records available for 25 years, and 4–9 year records available for 8 sites; a total of 53 sites (Hanson, 2001). The data for precipitation was subset to the interval 1 May 1974–30 May 1974 in order to capture the full spatial distribution of continuous records and a precipitation event that clearly registered on both the precipitation gauges and discharge measurements, thereby generating the best interpolated surface and minimizing the volume of data for the maximum number of sites (Figure 3). This temporal interval captures 49 of the precipitation measurement sites, excluding sites 138 × 22, 138 × 33, 138 × 44, 098 × 97 (see Hanson *et al.*, 2001 for site identification). Note that the precipitation follows the discharge slightly in Figure 3 as the nearest precipitation measurement site was to the southwest of the Weir that measured the discharge of Upper Sheep Creek (Site 138 × 31 in Figure 2), and the precipitation event moved in from the northeast.

METHODOLOGY

The methodology for the watershed model is based on the *nen* data model and implements a rule based approach to expressing the modelled processes. An advantage of a rule-based approach rather than equation-based is, easier inclusion of qualitative information, particularly for defining thresholds, such as expert 'non-encoded' knowledge (Seibert and McDonnell, 2000). The watershed runoff model aimed at capturing the dominant processes operating in a watershed while maintaining a level of simplicity that enabled the full development of the dynamic model in its breadth and the development of the simulator. The rules were derived from established specifications of these processes.

Model parameters

The watershed model included the following parameters: precipitation, elevation, bedrock, evapotranspiration, saturated hydraulic conductivity, infiltration capacity, and an initial water-table. Barring evapotranspiration, all of the parameters were created for the RCEW as a whole before being clipped to the Upper Sheep Creek watershed. The details of their creation are presented next.

Precipitation. Computed precipitation values were used for the dynamic model, and following restructuring of the data, a surface was interpolated for the whole of the RCEW from the 49 points of precipitation used. Each hourly set of measurements, within the temporal extent selected, were interpolated with universal kriging, which incorporated the DEM as the single trend component (Pardo-Iguzquiza, 1998).

Elevation. The relief of Upper Sheep Creek was modelled by a 30 m DEM provided in the RCEW dataset.

Bedrock. The soils data in the RCEW dataset included a field describing the depth to bedrock, which was used to generate a bedrock surface. However many of the values are unknown as they are deeper than those investigated. In these cases a value of $-x$ is given, meaning the bedrock is deeper than the specified value x. To solve this problem, 5 m was added to these absolute values. A bedrock layer was created by subtracting these depths from the DEM.

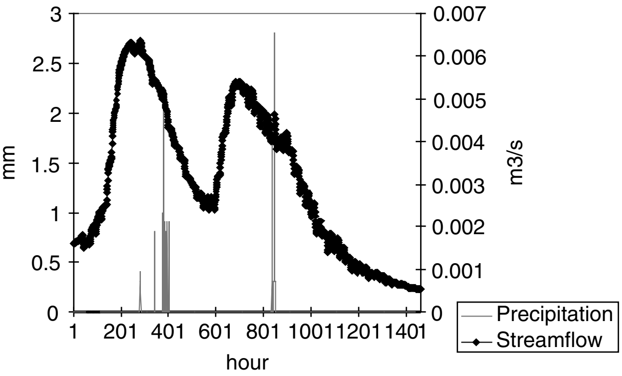

Figure 3. Precipitation, measured at site 138 × 31, and streamflow for Upper Sheep Creek sub-watershed from May 1974–June 1974

Evapotranspiration. Values for evapotranspiration in the Upper Sheep Creek were determined from a paper by Hanson and Wright (Hanson and Wight, 1995), providing a simple solution for defining evapotranspiration. They divided Upper Sheep Creek into two parts A and B, based on two types of vegetation, grass-low sagebrush and grass-mountain big sagebrush. On the basis of the vegetation layers, the two values of evapotranspiration were assigned to the different parts of Upper Sheep Creek (A and B). There is currently no diurnal variation expressed in the model, thus these evapotranspiration values are constant throughout the time period.

Saturated hydraulic conductivity. Using data from Rawls *et al.* (1982), definitions from the Soil Science Glossary provided by the Soil Science Society of America[2], and corresponding data in the RCEW dataset, saturated hydraulic conductivity was approximated from the soil texture class. Rawls *et al.* (1982) did not include silt in their categorization, thus it was approximated as being between silt loam and sandy clay loam, that is, as equal to 0·56 cm/h.

Infiltration capacity. The soil moisture data was measured at five sites, only three of which fall within the selected precipitation time frame. Three points are not enough to generate a surface therefore this data can only be used to calibrate or test the dynamic model. The Soil Hydrologic Group data was used to specify infiltration capacity, which is the National Resource Conservation Service classification for estimating overland flow[3]. As defined in the National Engineering Handbook (NEH-4), each hydraulic soil group is associated with an infiltration capacity. An added class was specified for the case where no infiltration could take place, such as on rock terrace escarpments.

Water-table. The water-table was created purely for the purposes of the model application, and is not expected to accurately represent the water-table in RCEW as it is unknown and the data are not available. The generated water-table has the streams within RCEW as its base, such that the cells at the location of perennial streams were assigned a value of 0 m below the DEM, and the intermittent stream cells were assigned a value of 1 m below the DEM. All other cells were assigned a value

---

[2] http://www.soils.org/sssagloss/tml.

[3] See the National Soil Survey Handbook produced by the National Resource Conservation Service (NRCS)—http://soils.usda.gov/technical/handbook/download.html.

based on an increasing function using the distance from these cells. This layer of values was then subtracted from the DEM, and the maximum value of this layer and the bedrock layer was taken as the water-table in order to ensure that the water-table was always above the bedrock layer.

Modelled processes specification

In what follows, an outline of the behaviour of each process represented within the dynamic model will be described in pseudo code. The processes included are infiltration, percolation, groundwater flow, Hortonian overland runoff, saturation excess runoff, and surface ponding. The global dynamics or emergent patterns that are generated from the simulation of the model result from the local interactions of these processes. The spatial extent of all processes is defined by the DEM, that is, by the selected rectangular area that represents the Upper Sheep Creek. The spatial granularity of each process is also defined by the DEM, where each process operates over a 30 m^2 area. The temporal extent of each process is defined by the model extent, that is, from 1 May 1974–30 May 1974. The temporal granularity of the processes is a function of the forcing parameters, which in the case of this dynamic model is the hourly update of precipitation.

For directing the processes that operated laterally in the x and y direction, such as runoff and groundwater flow, four routing algorithms were included in the model: D8, Rho8, FD8, and DInfinity (Freeman, 1991; Tarboton, 1997; Endreny and Wood, 2003). Following testing of the model under these routing scenarios, a mixture of DInfinity and D8 was used. D8 was only used when a threshold was exceeded that made DInfinity computationally untenable. Although the DEM and other parameters are defined at 30 m, the operation of these processes is defined on a continuous space and only their visualization (i.e. on a graphical user interface) is restricted to the 30 m by 30 m representation.

Following a precipitation event, processes are created as follows:

```
if (precipitation—evapotranspiration > infiltration capacity)
    if (there is a neighbouring point of lower elevation)
        create Hortonian overland flow
    else create surface ponding
else
    if (water-table is the same elevation as DEM)
        create saturation excess flow
    else create infiltration
```

Infiltration. The infiltration process converts directly to a percolation process at the following time step.

Percolation. Percolation processes are generated following infiltration and result in water flowing down in the z-direction through the soil matrix towards the water-table. The rate of downward flow is defined by the hydraulic conductivity parameter, and in the x- and y-direction according to the DEM surface.

```
if (the water-table has not been reached)
    if (there is a lower neighbouring elevation)
        percolate in a direction depending on
        the surface slope at a rate dependent on
        the hydraulic conductivity of the soil and
        the mass of water.
    else
        percolate straight down at a rate
        dependent on the hydraulic conductivity
        of the soil and the mass of water
else convert to groundwater flow
```

Groundwater flow. Groundwater flow occurs once percolation has reached the water-table.

```
if (there is a lower neighbouring cell based on water-table elevation)
    if (water-table >=DEM elevation)
        create saturation excess flow
    else
        continue flowing in direction of lowest
        water-table elevation
else
    if (water-table >=DEM elevation)
        create saturation excess flow
    else add to the water-table by elevating it
```

Hortonian overland runoff. Hortonian overland runoff is generated when the rate of precipitation exceeds the infiltration capacity of the soil.

```
if (Hortonian overland flow mass > infiltration capacity)
    if (there is a lower neighbouring cell in DEM)
        continue Hortonian overland runoff
        in direction of lowest neighbour
    else create surface ponding
else
    if (water-table >=DEM elevation)
        if (there is a lower neighbouring cell
        in DEM)
            create saturation excess flow
        else create surface ponding
    else create infiltration
```

Saturation excess runoff. Saturation excess runoff is generated when under precipitation, the water-table is equal to or exceeds the elevation of the DEM.

```
if (saturation excess mass > infiltration capacity OR water-table >=DEM elevation)
    if (there is a lower neighbouring cell in DEM)
        continue saturation excess runoff in
        direction of lowest neighbour
    else create surface ponding
else create infiltration
```

Surface ponding (SP). Surface ponding (SP) results when precipitation less evapotranspiration is greater than infiltration capacity and there is no neighbouring cell of lower elevation.

```
if (water-table elevation >=DEM elevation)
        continue surface ponding
else    create infiltration
```

Implementing the simulation environment

The simulation environment, called *flux*, was developed in Java using classes from the Repast (Recursive Porous Agent Simulation Toolkit) library, an open source agent-based modelling environment created by Social Science Research Computing at the University of Chicago (http://repast.sourceforge.net/). Repast is only used for its display and scheduling classes, and also has the advantage of containing Java classes for importing and exporting GIS raster data (ESRI ASCII raster files). Each of the processes described above are implemented as a class and flux iterates over these process classes, implementing the behaviour of their instantiated objects. The input parameters are also iterated over, which in this case is the precipitation that provides the forcing mechanism of the model.

Beyond the Repast tools for stepping through a model and visualizing the results, an extended GUI (graphical user interface) has been developed that allows for the selection of processes to be visualized (Figure 4). Figure 4 below provides an example of a simulation, where a visible chain of grey *nens* represent the process of groundwater flow in operation, underlain by a DEM with higher elevation represented by cells of lighter grey. This representation allows a user to explore the watershed for the location of modelled processes and better evaluate the efficacy of their model specification.

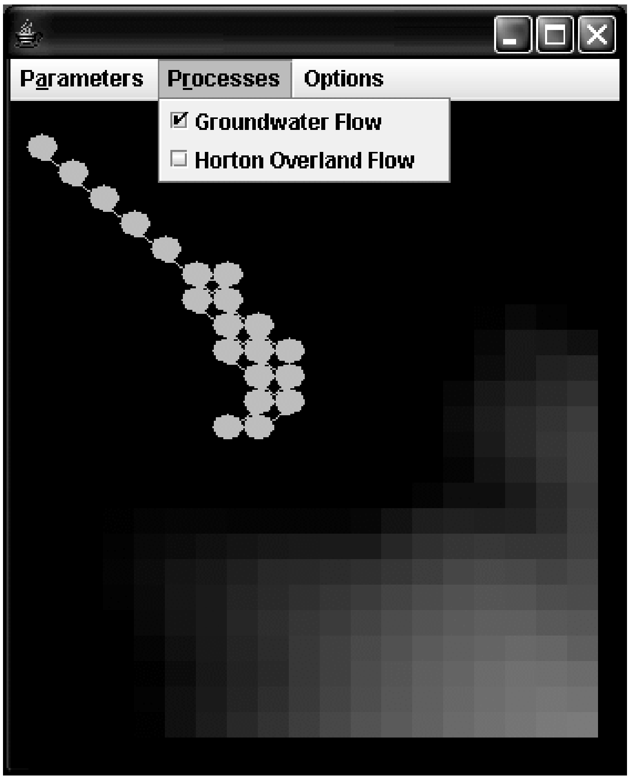

Figure 4. A sample graphical user interface (GUI) of the flux simulation environment

The results of a flux simulation are used to query for process information, namely for their state at an instant of time or their dynamics over an interval of time. These two base types of queries can be applied to properties or attributes of the processes, which includes spatial location. Given the nature of the process data model, the spatial character of a process includes: direction, location, and extent. A small tool was also developed in order to query the results of the simulation. For a full description of the simulation environment, see Reitsma and Albrecht (2005).

RESULTS

The hydrograph is the typical approach to analysing the results of a watershed runoff model, and it provides the initial means for exploring the results of the simulation presented in this paper. However, the use of the *nen* data model provides new scope for analysing the results of watershed simulations. In particular the spatial dynamics of the processes may be explored. In what follows, the initial results of a simulation, with the *nen* data model, is described.

Simulating a hydrograph is possible with the *nen* data model as it captures both state and process information. However, reproduction of watershed discharge over time is not difficult, nor does it imply that the processes in the dynamic model have been adequately modelled (Beven, 2000). Figure 5 below provides a hydrograph of the simulation results below the measured hydrograph. The values of the y axis on the simulated hydrograph provide a relative measure of discharge that directly corresponds

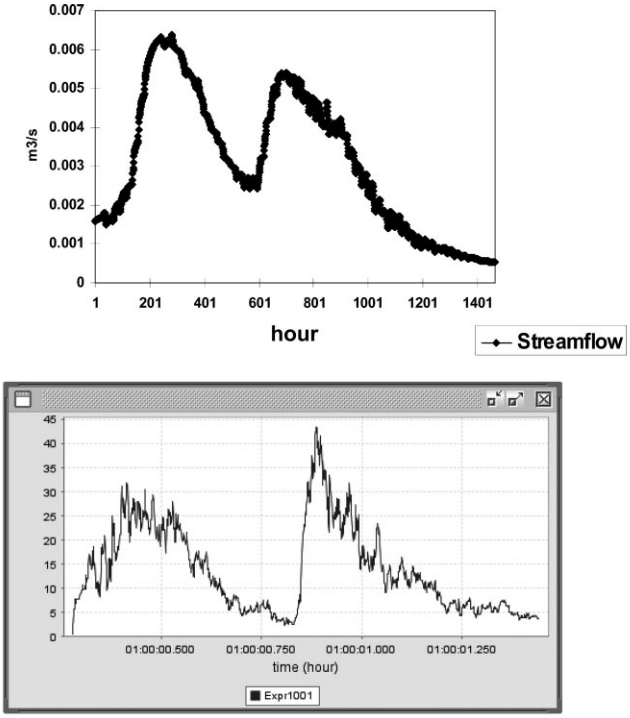

Figure 5. Empirical hydrograph (top), and simulated hydrograph (bottom) of Upper Sheep Creek 1 May 1974–30 May 1974

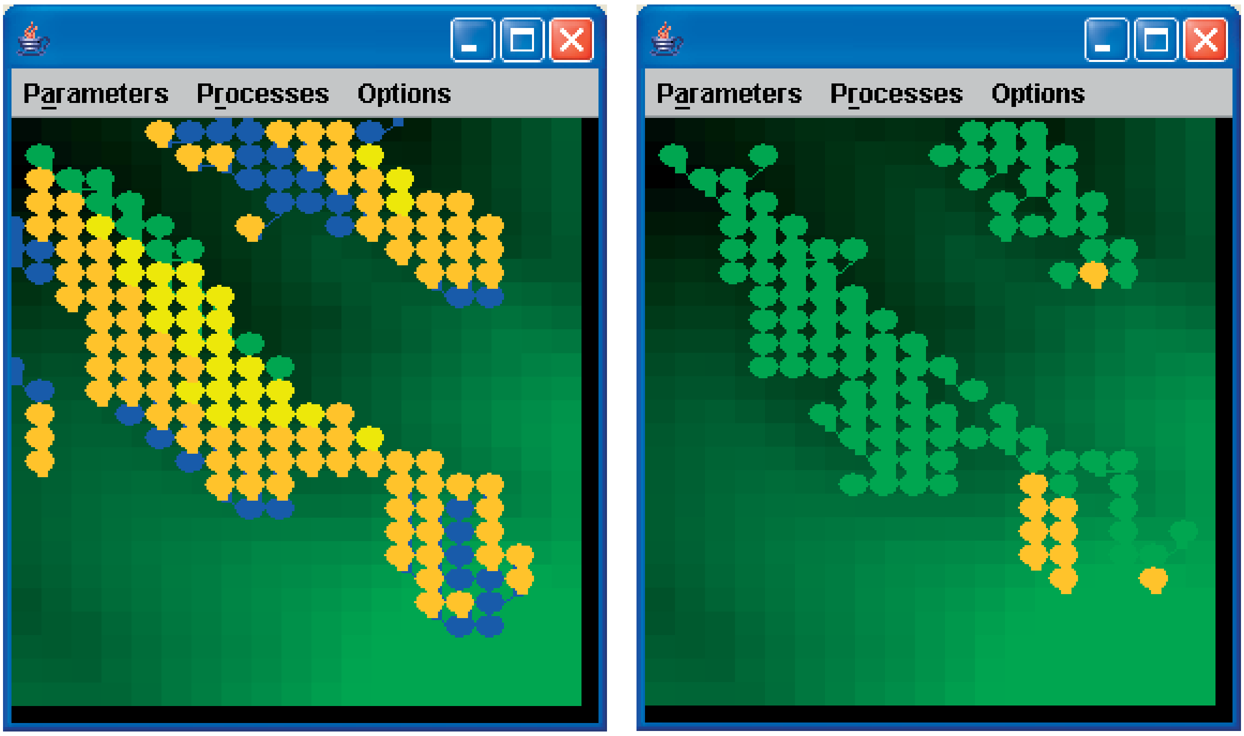

Figure 6. Simulation display output for hour 402 (left) and hour 412 (right)

to the precipitation input given in mm. The simulated hydrograph (Figure 5, bottom) is reasonably consistent with the observed hydrograph (Figure 5, top). The two primary peaks and the general patterns of flow are captured. Beyond any problems with the specifications of the processes, such as increased saturation impacting on other runoff processes, the primary sources of discrepancy include the model not taking into account baseflow that results from spring snowmelt, and the 30 m × 30 m grid scale used for the operation of some of the processes, which causes the fluctuations in the results.

The advantage of the process data model, however, is that we can move beyond the hydrograph as our main form of validation and start to explore how the processes defined in our model are expressed at runtime, and what their spatial, temporal, and attribute characteristics are. The first result to consider is the spatial dynamics of the modelled processes. Figure 6 below presents two process time slices, displaying the spatial distribution of the processes at hour 402 and hour 412 over the DEM. The green *nens* represent groundwater flow, blue *nens*: Hortonian overland runoff, orange *nens*: percolation, and yellow *nens*: infiltration. This display allows users to compare their process descriptions in the model with qualitative knowledge of where those processes occur in reality. Unfortunately, in the case of the Upper Sheep Creek sub-watershed such detailed descriptions of watershed processes are not available in the literature, and thus validation of the process definitions is problematic. This is not a shortcoming of the present dynamic model, but rather highlights a potential evolution of how validation must be approached when using nen-based implementations, as discussed below.

Beyond the spatial qualities of the results, any other aspect of the process that is stored in the data structure may be queried and analysed. For example, the query tool created for the *flux* modelling framework also allows for direction based querying; however, this would perhaps be more useful for other types of processes such as atmospheric processes. Although not yet implemented in the tools developed for this research, in a model that incorporates the interaction of processes at different scales, representing the process information will allow for novel queries such as selecting spatio-temporally coincident or interacting processes or for tracing the dynamics of individual process instances. The results of such queries will form the basis for further analysis.

DISCUSSION

Simulating a dynamic model with the *nen* data model may provide new insights into where processes dominate in the modelled watershed system, about how those processes interact, and give modelers the opportunity to evaluate how the spatial dynamics of their process descriptions match what they expect to see in the real world. While other simulation approaches may isolate the results of certain processes, without a data model to represent those processes we cannot easily visualize or analyse them at a time or as they change over time. By explicitly representing the process information that is implicit within dynamic models, we can visualize and analyse how they operate and interact with other processes, supporting explanatory models that seek to accurately capture the real world they reflect rather than predictive models that may focus on associating the right input with the appropriate output. For example, simulating a watershed model with the *nen* data model has allowed us to determine whether the spatial dynamics of the process, defined by the rules governing their behaviour, reflects what we expect to see out in the field.

The methodology also lends itself to model intercomparison studies, such that the distribution and quantity of process instances can be compared among dynamic models. This is in contrast to traditional approaches that analyse the state of the modelled system at the end of the simulation or over consecutive time steps, using this information to compare models and evaluate their accuracy (for example Dutay *et al.*, 2002). Rather, with an appropriate data model and formal definitions of the process' characteristics, the process dynamics that are specified in the mathematical formalisms of the dynamic model can be represented, showing the spatial distribution and the temporal dynamics of those processes. This then allows us to automatically express the qualitative nature of the output of the model, and associate that output with extant information about those processes available in knowledge bases or utilize that output in new and interesting ways. Representing the processes in operation at each time step enhances our ability to understand the results of the dynamic model and evaluate our mathematical formalisms used to model processes.

At this stage, the *flux* simulation environment is constrained to small models due to limitations of computational complexity. The scaling of the environment to large watersheds with more processes and detail, is theoretically straightforward, but computationally intensive. As an alternative to extending the *flux* environment, it should not be difficult to modify existing modelling software environments to implement the *nen* data model. Two aspects of the software would need to be changed, namely the code that records the process in action and the code that displays the results of the model.

Another limitation of using this approach is in validating the results. Again, this is not a limitation of the *nen* data model, *per se*, but rather of the typical way validation is performed. Validation occurs by matching the output of the model with the real world, a good result being the ability to mirror that world *in silico*. Endreny and Wood (2003), for example, qualitatively validate their simulated flow networks with empirical data. The standard approach to validation in watershed modelling is to compare the simulated output of volume of stream discharge over time, with discharge measurements over the same time period for the modelled watershed. More recently, validation approaches have attempted to use remotely sensed data to observe the spatio-temporal distribution of surface state variables that are important for or in some way diagnostic of processes, and compare these with model outputs. For example, many land surface hydrology models predict surface temperature, which can be compared with satellite-derived estimates of surface temperature. The temperature, in turn, plays an important role in evapotranspiration, which then is traced eventually to the water balance. Such indirect methods have had some success.

However, in validating and fully testing the dynamic model results described above using the *nen* data model, the central problem is that long-term empirical observations of the processes themselves (as opposed to the surface state variables) are not available for describing their location and duration. Some of the literature covering RCEW does provide limited discussion on the processes operating in certain parts of the watershed, however, this is not enough for model validation. Without such real world data, any dynamic model developed with the process-based methodology described in this paper cannot be effectively validated. This can be defined as a form of process modelling equifinality, where the same system state can result from many different process pathways, which is well recognized by watershed modelers as a problem of validating against hydrographs (Beven, 2000). A possible solution would be to validate the model against another model of similar nature, yet no such model exists. As such, validation of a fully specified domain model is left to a future research objective. This will involve intensive study of a particular watershed and the development of appropriate measurement methods that either standardize qualitative descriptions or propose new process based measurement approaches.

The results presented above are promising in their representation of watershed processes but can be further explored through the development of new analytical measures. The *nen* data model provides new scope for visualization and analysis of spatial processes, which is a goal of our continuing research. Applied to watershed modelling, these analytical techniques will inform those studying the model results for process patterns and could provide a new approach to characterizing or classifying watersheds, as suggested by Vogt *et al.* (2003), who consider the thresholds of hillslope processes as defining the hillslope-valley transition. For example, we might classify watersheds based on a set of dominant processes. Furthermore, because the *nen* is conceptually based on a continuous notion of space and time, it is not constrained to regular spatial or temporal tessellations, for example, TINs may be used as an underlying spatial representation for an input parameter, or some other event based notion of time.

Given the potential of data models to shed new light on simulations, new open and flexible modelling platforms are needed that can easily accommodate new data models and new analytical and visualization techniques. Such a platform would provide a unique scientific environment for not only testing new models but testing new simulation methodologies. This vision for the future of watershed modelling and dynamic modelling in general goes far beyond the flux prototype discussed in this paper, but provides direction for its future development and for simulating larger scale watershed systems.

CONCLUSION

Simulating a watershed with the *nen* data model provides new insights into the results of a dynamic model. Current methods for simulating watershed runoff have focused on providing the state of the modelled system as their results, states describing the location of matter or attributes of the system. The general methodology described in this paper presents a complementary view of the modelled system, and has the advantage of being able to extract the information about the state of the modelled system in terms of the distribution of matter such as water and soil. Its novelty is the provision of a new epistemological window on the modelled results, allowing for new, process-oriented queries and analysis. The hydrograph results of the simulation provided fairly good agreement with the observed results of runoff in Uppper Sheep Creek over the same time period. However, validation of the watershed model requires new methods for collecting information about the processes operating in the watershed.

One of the important motivations of the methodology presented here is to provide an environment that facilitates exploratory and explanatory insights of model results, rather than solely for predictive purposes. However, we believe that *nen*-based modelling approaches with their focus on process may ultimately represent physical systems in a way that generates the best prediction of its future state. The goal is to represent the system *in silico* so that we can expect changes in our dynamic model to reflect changes in the real world.

REFERENCES

Abbott MB, Bathurst JC, Cunge JA, O'Connell PE, Rasmussen J. 1986a. An introduction to the European Hydrological System—Systeme Hydrologique Europeen, "SHE", 1: history and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology* **87**: 45–59.

Abbott MB, Bathurst JC, Cunge JA, O'Connell PE, Rasmussen J. 1986b. An Introduction to the European Hydrological System—Systeme Hydrologique Europeen, "SHE", 2: structure of a physically-based, distributed modelling system. *Journal of Hydrology* **87**: 61–77.

Baird A. 2004. Soil and hillslope hydrology. In *Environmental Modelling: Finding Simplicity in Complexity*, Wainright J, Mulligan M (eds). John Wiley and Sons: Chichester.

Beven KJ. 2000. Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences* **4**(2): 203–213.

Beven KJ. 2001. *Rainfall-runoff Modelling: The Primer*. Wiley: Chichester.

Beven KJ. 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes* **16**: 189–206.

Borah DK, Bera M. 2003. Watershed-scale hydrologica and nonpoint-source pollution models: review of mathematical bases. *Transactions of the American Society of Agricultural Engineers* **46**(6): 1553–1566.

Borah DK, Bera M. 2004. Watershed-scale hydrologica and nonpoint-source pollution models: review of applications. *Transactions of the American Society of Agricultural Engineers* **47**(3): 789–803.

Coulthard TJ, Kirkby MJ, Mackline MG. 2000. Modelling geomorphic response to environmental change in an upland catchment. *Hydrological Processes* **14**: 2031–2045.

De Roo APJ, Wesseling CG, van Deursen WPA. 2000. Physically-based river basin modelling within a GIS: the LISFLOOD model. *Hydrological Processes* **14**: 1981–1992.

Dutay JC, Bullister JL, Doney SC, Orr JC, Najjar R, Caldeira K, Campin JM, Drange H, Follows M, Gao Y, Gruber N, Hecht MW, Ishida A, Joos F, Lindsay K, Madec G, Maier-Reimer E, Marshall JC, Matear RJ, Monfray P, Mouchet A, Plattner GK, Sarmiento J, Schlitzer R, Slater R, Totterdell IJ, Weirig MF, Yamanaka Y, Yool A. 2002. Evaluation of Ocean Model Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models Ventilation with CFC-11: comparison of 13 Global Ocean Models. *Ocean Modelling* **4**: 89–120.

Endreny TA, Wood EF. 2003. Maximizing spatial congruence of observed and DEM-delineated overland flow networks. *International Journal of Geographical Information Science* **17**(7): 699–713.

Favis-Mortlock DT, Boardman J, Parsons AJ, Lascelles B. 2000. Emergence and Erosion: a model for rill initiation and development. *Hydrological Processes* **14**: 2173–2205.

Feng C, Sorokine A. 2001. Incorporating hydrologic semantic information for interoperable GIS with hydrologic model. In *Paper Read at Proceedings of the 9th ACM International Symposium on Advances in Geographic Information Systems, at Atlanta*, Georgia.

Freeman TG. 1991. Calculating catchment area with divergent flow based on a regular grid. *Computers & Geosciences* **17**(3): 413–422.

Freeze RA, Harlan RL. 1969. Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology* **9**: 237–258.

Gao X, Sorooshian S, Goodrich DC. 1993. Linkage of a GIS to a distributed rainfall-runoff model. In *Environmental Modeling with GIS*, Goodchild MF, Parks BO, Steyaert LT (eds). Oxford University Press: Oxford; 182–187.

Grayson R, Blöschl G. 2000. Spatial modeling of catchment dynamics. In *Spatial Patterns in Catchment Hydrology*, Grayson R, Blöschl G (eds). Cambridge University Press: Cambridge; 51–81.

Gurtz J, Zappa M, Karsten J, Herbert L, Verbunt M, Badoux A, Vitvar T. 2003. A comparative study in modelling runoff and its components in two mountainous catchments. *Hydrological Processes* **17**: 297–311.

Haff PK. 2001. Waterbots. In *Landscape Erosion and Evolution Modeling*, Harmon RS, Doe WW (eds). Kluwer: New York; 239–275.

Hanson CL. 2001. Long-term precipitation database, reynolds creek experimental watershed, Idaho, United States. *Water Resources Research* **37**(11): 2831–2834.

Hanson CL, Marks D, Van Vactor SS. 2001. Long-term climate database, reynolds creek experimental watershed, Idaho, United States. *Water Resources Research* **37**(11): 2839–2841.

Hanson CL, Wight JR. 1995. Evapotranspiration fluxes within a rangeland watershed. In *Paper Read at Fifth International Rangeland Congress*, at Salt Lake City, Utah.

Maidment DR. 1993. GIS and hydrologic modeling. In *Environmental Modeling and GIS*, Goodchild MF, Parks BO, Steyaert LT (eds). Oxford University Press: Oxford; 147–167.

Marks D. 2001. Introduction to special section: reynolds creek experimental watershed. *Water Resources Research* **37**(11): 2817.

Miller SN, Semmens DJ, Miller RC, Hernandez M, Goodrich DC, Miller WP, Kepner WG, Ebert D. 2002. GIS-based hydrological modeling: the automated geospatial watershed assessment tool. In *Paper Read at Proceedings of the Second Interagency Hydrologic Modeling Conference*, July 28-August 1 2002, Las Vegas, Nevada.

Mulligan M. 2004. Modelling catchment hydrology. In *Environmental Modelling: Finding Simplicity in Complexity*, Wainright J, Mulligan M (eds). John Wiley and Sons Ltd: Chichester.

Pardo-Iguzquiza E. 1998. Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. *International Journal of Climatology* **18**: 1031–1047.

Pullar D. 2003. Simulation modelling applied to runoff modelling using mapscript. *Transactions in GIS* **7**(2): 267–283.

Rawls WJ, Brakensiek DL, Saxton KE. 1982. Estimation of soil water properties. *Transactions of the American Society of Agricultural Engineers* **25**(5): 1316–1320.

Reitsma F, Albrecht J. 2005. Implementing a new data model for simulating processes. *International Journal of Geographical Information Science* **19**(10): 1073–1090.

Romanowicz R, Beven KJ. 1993. GIS and distributed hydrological models. In *Geographical Information Handling—Research and Applicaitons*, Mather PM (ed.). John Wiley and Sons: London; 197–205.

Seibert J, McDonnell J. 2000. Towards a better process representation of catchment hydrology in conceptual runoff modelling. In *Paper Read at IAHS Conference on Meso-scale Modelling. Proceedings of* *the International Workshop on Runoff Generation and Implications for River Basin Modeling*, 9–12 October 2000, Freiburg.

Seyfried M, Harris R, Marks D, Jacob B. 2001a. Geographic database, reynolds creek experimental watershed, Idaho, United States. *Water Resources Research* **37**(11): 2825–2829.

Singh VP (ed.). 1995. *Computer Models of Watershed Hydrology*. Water Resources Publications: Highlands Ranch, CO.

Slaughter CW, Marks D, Flerchinger GN, Van Vactor SS, Burgess M. 2001. Thirty-five years of research data collection at the reynolds creek experimental watershed, Idaho, United States. *Water Resources Research* **37**(11): 2819–2823.

Streit U, Wiesmann K. 1996. Problems of integrating GIS and hydrological models. In *Spatial Analytical Perspectives on GIS in the Environmental and Socio-Economic Sciences*, Fischer M, Scholten H, Unwin D (eds). Taylor and Francis: London; 161–173.

Tarboton DG. 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research* **33**(2): 309–319.

van Oosterom P, Maessen B, Quak W. 2002. Generic query tool for spatio-temporal data. *International Journal of Geographical Information Science* **16**(8): 713–748.

Veith TL, Srinivasan MS, Gburek WJ. 2003. *Process representation in watershed-scale hydrologic models: an evaluation in an experimental watershed*. In *Paper Read at First Interagency Conference on Research in the Watersheds*, October 27–30 2003, Benson, AZ.

Vogt JV, Colombo R, Bertolo F. 2003. Deriving drainage networks and catchment boundaries: a new methodology combining digital elevation data and environmental characteristics. *Geomorphology* **53**(3–4): 281.

Ward RC, Robinson M. 2000. *Principles of Hydrology*. McGraw-Hill Publishing Company: Maidenhead.