Available online at www.sciencedirect.com

SCIENCE DIRECT®

Environmental Modelling & Software 21 (2006) 1340–1352

Environmental Modelling & Software

www.elsevier.com/locate/envsoft

Hydrological models are so good, do we still need data?

R.P. Silberstein*

CSIRO Land and Water, Private Bag No. 5, PO Wembley, WA 6913, Australia

Received 4 August 2004; received in revised form 6 April 2005; accepted 6 April 2005
Available online 5 October 2005

Abstract

Our ability to numerically model natural systems has progressed enormously over the last 10–20 years. During the last decade computational power has increased to the stage where we can now have a super-computer on our desk, and the detail and fine scale processes that can be included in models are fantastic. The computational tools available for analysis and display have opened doors beyond the dreams of our forebears. However, as modelling power has increased there has been a concurrent reduction in "data power", particularly in the collection of hydrological data. While we undoubtedly have access to large datasets of extraordinary technological finesse such as the remotely sensed data from satellites, our collection of more basic and traditional datasets suffers. We can read car number plates from outer space, but we still, in the main, measure rainfall by putting a bucket out in a paddock. This paper discusses the growth in sophistication of hydrological modelling through the last hundred years. The concept of validation or verification of models is questioned, and the role of data in modelling discussed. It is argued that modelling in the absence of adequate data is not science, unless it is to develop hypotheses that are to be tested by observation. Several modelling case studies with and without adequate testing data are discussed. It is also argued that improvement in the management of our environment and water resources will not come with improved models in the absence of improved data collection because we cannot manage what we do not measure.

Keywords: Modelling; Hydrology; Monitoring; Data collection; TOPOG; LASCAM; IQQM; TOPMODEL

1. Introduction

Modelling is now a common tool in many fields of scientific endeavour. Physical scale models have been used to study the static and dynamic behaviour by engineers for many years, to ensure their bridges would stay up, or their breakwaters would not wash away. Children construct scale models of racing cars and aeroplanes to explore the effect of high speed crashes and glide angles of fighters without engine power.

Mathematical modelling is generally not visually impressive but it allows exploration of real behaviour that would be difficult to observe or measure in nature. Mathematical models are used in economics, politics and policy, finance, commerce and the behavioural sciences. Modelling these systems is generally impossible without simplifying the representation of the real system simulated. Statistical and behavioural models are used to predict horse races, stock markets, ecological systems and social systems. These models are used to try to predict how the systems will develop over time, usually so that some return may be maximised for the modeller – this may be financial, political return, or in order to improve social harmony.

For scientists, the aim is to construct a model representing some component of the real world. For some, the model development is sufficient in its own right, such as Game Theoreticians or economists, who explore the impact of various rules on system outcomes. These

* Fax: +61 8 9333 6211.
E-mail address: richard.silberstein@csiro.au

1364-8152/$ - see front matter
doi:10.1016/j.envsoft.2005.04.019

activities may be confused with real world studies because of the common terminology that has crept into much economic and political discourse, and the techniques are used to study real world situations, but in the scientific sense are mainly mathematical games.

Science is founded on observations. Indeed, science is defined as "the systematic study of man (and woman) in their environment based on the deductions and inferences that can be made from reproducible observations and measurements of events and parameters within the universe" (The Macquarie Dictionary, 1987). Data are science, models are a complement to them, but not a replacement for them.

A model is, by definition, a simpler representation of the real thing. It is essentially a toy, albeit a useful one, as a mathematical mimic of the real, more complicated, system. It is not a unique opinion that modelling is fine as long as it is not confused with the real thing.

This paper was stimulated by comments received when the author was attempting to reinvigorate hydrological monitoring in Australia. While senior bureaucrats in data collection agencies and their policy clients generally appreciate the need to collect data on the systems they are trying to manage, two comments made were that "models have reached a level of sophistication that renders data collection less important" and "data collection these days is principally to calibrate models". These comments strike the author as presumptuous and ignorant, and perhaps arrogant, and this paper is an attempt to place the dual activities of modelling and field data collection in some relative context.

2. Hydrological modelling

While models of the water cycle have been around since at least the Ancient Greeks, hydrological modelling in a mathematical sense probably began when M. Darcy (1856) published his analysis that water flows down a pressure gradient, at some rate determined by a property of the medium (hydraulic conductivity) through which it is flowing. This was analogous to Ohm's Law, and indeed electric circuits have been used as analogue models of hydraulic systems for much of the period since.

Saint-Venant (1871) gave a mathematical description of river hydraulics and Richards (1931) significantly improved Darcy's "Law" by adding conservation of mass and unsaturated dispersion. Horton (1933) provided a landmark in the description of runoff generation, and it was then theoretically possible to model a catchment with mathematical representations of infiltration, groundwater flow, surface runoff generation and river routing. The mathematical description of runoff was significantly improved with the addition of kinematic surface flow (Brakensiek, 1967), and thus by the 1970s it was possible to model the basic water pathways once rainfall strikes the ground. Hydrological modelling has since progressed to the stage where virtually any catchment process can be included.

The single biggest advance for modelling the reverse flux – evaporation – came with Penman (1948) model based on data collected during World War II, elegantly added to by Monteith (1965). Vegetation–water interactions have been added: rainfall interception (Rutter et al., 1971; Gash et al., 1980), transpiration, root-water uptake (for example, Ritchie, 1972; Ritchie and Otter-Nacke, 1985) and growth (for example, Wu et al., 1994). The representation of vegetation now includes the dynamic response to environmental conditions, with carbon assimilation and partitioning above and below ground in response to stresses in one dimension (Dawes et al., 1998) and in full catchment models (Silberstein et al., 1999b; Vertessy et al., 1996). Thus we have methods to simulate runoff generation, subsurface unsaturated and saturated flow, river storage and routing and evaporation and transpiration.

Mathematical modelling has now become a whole field in itself, with dedicated textbooks, courses, and conferences (see for example, Fowkes and Mahony, 1994; the series of MODSIM conferences, and this journal). I suggest that in many cases models take on a life, especially in the eyes of their creators, that equals or exceeds their relevance to the real world. Whole conferences have been devoted to single models (such as the TOPMODEL workshop in 1995, see Beven, 1997), and major experiments established investigating, not the real world but inter-model comparisons, as if they were different races of a new species (e.g., Project for the Intercomparison of Land Surface Parameterization Schemes – PILPS, Henderson-Sellers et al., 1993; Pitman and Henderson-Sellers, 1998).

Over the last 30 years, I suggest, hydrological science has divided with two approaches used independently by the *modellers* and the *experimentalists*, and often the two groups having little to do with each other. Occasionally, modellers may stray well out of their territory and seek data for input, or even more rarely data for validation of their model, but this is still unusual, although thankfully becoming less so. It is my view that modellers who focus on their model without continual reference to real data are not really scientists but artists. They have their place, and indeed their ideas may well turn out to be useful, but their activities are not science if they are not based on observations.

3. Hydrological model species

Beven (2001) has given an excellent expose on the art or science behind hydrological modelling, and I will not attempt to cover his material, however, some introduction to modelling approaches is required. Modelling a catchment involves a decision of fundamental philosophy as to whether the model should be "lumped" or "distributed", and whether it should be "deterministic" or "stochastic" (Beven, 2001). "Lumped" models represent a catchment as a single entity, or a small group of entities, such as reservoirs, and simulate state variables and fluxes into and out of the catchment as a whole. "Distributed" models divide the catchment up into many entities, each representing small parts of the catchment, and the state variables and fluxes between the entities are determined across the catchment. There are also models that fall somewhere between these, in which variables are not explicitly distributed across the catchment but are represented as a distribution and could therefore be interpreted anywhere on the catchment (for example, TOPMODEL in Beven et al., 1995 and Beven, 1997; MACAQUE in Watson et al., 1999).

Beven (2001) also discusses the choice between "deterministic" and "stochastic" models. Simulations with "deterministic" models will always produce the same answer if they have the same input data, whereas "stochastic" models will result in a distribution of results or a result with an accompanying variance. There are also many examples of using deterministic models in a stochastic way, to generate a distribution of results from many simulations with a distribution of input parameters (the so-called Monte-Carlo method).

Whether distributed or lumped, hydrological, and other, models fall into three main types, such that each component is either "conceptual", "empirical" or "statistical", or "physical" (Wheater et al., 1993; Grayson and Chiew, 1994). The vast majority of models in the literature are "conceptual" or "stocks and flows" models; in hydrological terms a catchment is represented as a series of moisture stores, with fluxes between the stores and out of the catchment represented by parametric equations. The stores and fluxes are usually identifiable with real properties and processes, but are not generally independently measurable; the parameters must be estimated from input and output data (Ye et al., 1997).

Empirical models are generally the simplest models and are utterly dependent on data as they represent the relationship between input and output series, generally as "transfer functions" between these series. Ye et al. (1997) discussed the performance of conceptual and empirical models in simulating the streamflow of semi-arid catchments in Australia. This study was one of a number of inter-model comparisons that are absolutely dependent on data, not only for their tests but the models themselves could not be developed without calibration data. Indeed, there is no reason to suppose that things should be otherwise.

At the lowest level of physical complexity are the simple "bucket" conceptual models, such as those presented by Boughton (1995). In these models, the whole catchment is represented by a single or a small number of buckets. The underlying principle is that a catchment has two major properties that control most of its response to drivers – the ability to store water and the rate of release of that water. Depending on the storage size, relative to the timescale and intensity of forcing, streamflow is generated. This idea was explored in some detail by Farmer et al. (2003) who showed that most catchments could be well represented with a small number of buckets, and the exact number depended on relatively few characteristics.

In the last five years or so personal computers have developed enough power that very sophisticated distributed models can be run on moderately sized catchments. In fact, there is a reasonable argument that we would not have developed these sophisticated models without the power to exploit it. At the same time as these advances, we have seen the development of much simpler models. For example, TOPMODEL (Beven and Kirkby, 1979) uses a relatively simple terrain based attribute (the "topographic wetness index") to drive catchment flow and water storage processes. In doing so it retains some of the internal complexity of catchments but is much simpler than the more complete physical representations.

At a similar level of complexity, but using a different approach, LASCAM (Sivapalan et al., 1996a,b) represents a large catchment as a manageable number of sub-catchments, each represented as a single lumped entity. Most of the model parameters are assumed to be the same for each of the sub-catchments, and are determined by calibration on streamflow response to rainfall and potential evaporation drivers.

Hybrid models (such as Silberstein et al., 2003b), combine relatively simple conceptual catchment models (like LASCAM) to more complicated energy balance models to determine evaporation and surface temperatures. The surface temperatures can then be linked to satellite data, either for model testing (Silberstein et al., 1999a) or as inputs (McVicar and Jupp, 1999).

At the most complicated end of the model spectrum, physically based hydrological models may, in principle, be operated without streamflow data for calibration, as they purport to represent the important physical processes with parameters that can be measured independently, and assigned a priori to the relevant model characteristics (e.g. *SHE*, Abbott et al., 1986; *IHDM*, Beven et al., 1987; *TOPOG*, Vertessy et al., 1996). In the quest for truth and beauty these models have become so sophisticated that their successors would seek to include moisture, solute and suspension fluxes at minute levels of detail. In practice, however, this approach is defeated by the lack of sufficient data to adequately characterise the model, by the fact that no model represents all the internal heterogeneity, and that they do not have means of including processes internal to the catchment unit (Grayson et al., 1992). Thus, the most sophisticated of catchment models still rely on input and output data for some level of calibration because there is never enough characterisation to avoid it.

These sophisticated models give tremendous tools for investigation, but of what? If they are used as engineering tools, the boundary conditions and the domain must be clearly defined and valid. If they are used for policy development it is essential to represent their uncertainties, and as scientific tools they should be used to test hypotheses.

The ultimate in model computational sophistication is probably represented by the global climate models, which are run continuously on the world's most powerful computers, and are linked to "meso-scale" meteorological forecasting models. We now have the capacity to simulate the climate over the last hundreds of years, and the next, and produce streamflow for water resource managers for decades, or centuries, to come. These capabilities are truly awesome, and genuinely raise the question: Given we can now simulate how the world works so well, can we cut down the cost of collecting data that are perhaps no longer needed?

All models have spatial and temporal limits to their discretisation and description, which is another way of saying that the "scale problem" remains unsolved. While the physically based models cannot be run seriously without data, the empirical and conceptual models cannot really be run *at all* without data.

4. Model uncertainty

Uncertainty in modelling can arise from errors in model structure and in parameter estimation. Assuming the right structure, models may be improved by adding processes, and hence complexity, requiring more parameters. Thus structural uncertainty tends to decrease (in a good model), but parameter uncertainty may increase. Fig. 1 illustrates this (following Passioura, 1996; Reynolds and Acock, 1985), with the left image illustrating a well structured model that fits reality, with cumulative errors growing as the number of parameters grows, but the structural errors diminishing with this extra complexity. If the structure is flawed (image on the right), no amount of increasing parameters will reduce this error, and the model can only reach its structural accuracy, not its parametric accuracy. In this case, there is an opportunity for scientific progress because the only way of reducing the errors is by improving the model; thus we may learn when models fail to reproduce reality, but this is only possible because we have real data with which to compare our model. Passioura (1996) cites a number of examples in which flawed models resulted in fundamental changes in understanding the way plants grow. Perhaps the most dangerous thing in hydrology may be a model that fits with expectations (Dooge, 1988), because then, if we accept that our encapsulated understanding is adequate, we are not progressing.

Classic model structural errors were demonstrated when Michelson and Morley failed to measure the speed of Earth's travel through the interstellar "ether" and ended up determining that the speed of light was the same in all directions. Einstein examined the differences between Newton's Law of Gravity and observations of planetary movements and ended up developing relativity – both "special" and "general". Both of these results appeared esoteric, but have huge implications on our understanding of the universe and our place in it, and both were data driven resulting in fundamental changes to our model of the Universe. They also have fundamental repercussions on several aspects of modern life; nuclear energy released in bombs and power stations was first calculated by Einstein as part of Special Relativity, and gravitational corrections are required for the now ubiquitous geographical positioning system (GPS) used for navigation. In contrast, Einstein won his Nobel Prize in Physics for a model developed in the absence of data, the theoretical prediction of the photoelectric effect, several years before there were any observations to support it. Thus, there clearly is a role for pure thought, theory,

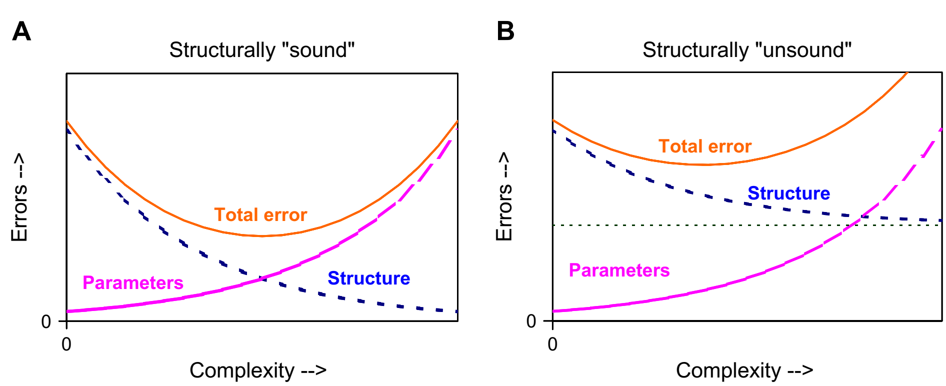

Fig. 1. Notional components of model prediction error. (A) Structurally sound model, with a tendency to a global minimum in prediction error arising with the optimum combination of number of parameters and complexity. (B) Structurally unsound model with a higher prediction error level that does not reduce as complexity is added (after Passioura, 1996, and Reynolds and Acock, 1985).

and mathematical games, but they only become useful if and when they are tested with measurements.

Part of the problem of determining structural errors is demonstrated when a model does reproduce observations. The problem then is that it is not known whether this is because the model is right, or the errors cancel out, and it is certainly not known whether the errors are structural or observational. This is part of the problem of non-uniqueness of models (Beven, 2000). Two models may be empirically equivalent if there is no way to distinguish between them without invoking some external subjective factor, such as simplicity, symmetry, elegance or simply personal preference (Oreskes et al., 1994). For example, the notions of a flat or spherical Earth were indistinguishable until new evidence was obtained by Magellan and others.

5. The use of models

Models serve three main purposes. Firstly, they give us a framework to assemble our process understanding and to explore the implied system behaviours that come from that understanding. We can examine the model results, and consider whether they concur with our overall system analysis or not. If not, we have a structured framework to analyse whether it is our model or our overall understanding, or both, that is in error.

The second main use of a model is as a mechanism for testing data, to check for inconsistencies and errors, and to fill in missing information. It also gives us a method to explore the implications of our measurements. In fact, this may be the most useful function of models, because they help structure scientific enquiry that can elucidate further details behind observations. These first two uses of models fall into the categories of *scientific* and *engineering* discussed by Passioura (1996). If we intend using models as scientific tools, we need to ensure that we actually use them to test hypotheses, and not just play computer games that reinforce our understanding and the conjectures built into the models.

The engineer will take a model to solve some day to day problem within well defined boundary conditions for which the responses are well understood, and the model essentially provides a decision support system for them and their client. It gives a semblance of authority and a legally defendable recommendation. However, if the engineer strays outside the well defined parameter boundaries into "extrapolation space", they are entering the *gaming domain* which may well be dangerous.

The third use of models, and probably the most widely publicised and "commercial" use, is to explore scenario options. These may be options for management of a system or exploring possible outcomes under a range of different input conditions, perhaps depending on future climate, political or economic scenarios. However, unless these scenarios are well constrained within known data boundaries, it is my contention that these activities should be confined largely to stimulate discussion, should always be tempered by some healthy scepticism, and retain due regard for our understanding of the whole system being considered. At their worst, in a scientific sense, these activities are simply computer games, that Passioura (1996) likened to snake oil, and those that report their results as the salesmen of the American wild west. Passioura argued that the highly sophisticated models "at best, give structural insights to their developers, and at worst are merely time-wasting ceremony". Modelling scenarios definitely have a place in planning and management decisions, but there needs to be a healthy reality check on the processes. Models are used because they are much cheaper and faster than doing real experiments. They also have the ability to predict things that we may not be able to do in the real world, or perhaps that would not really happen.

5.1. Scenario modelling examples

The models are also used because they often provide a well organised set of output that can be used to generate graphs and images representing the possible outcomes. The results of these modelling exercises are used to inform political and environmental decision making. Models like the Integrated Quantity and Quality Model (IQQM), developed by the New South Wales Department of Land and Water Conservation (DLWC, 1995), can simulate water quality at any point along the 3000 km long Murray River. Viney et al. (2003) studied a small section. With simple tools graphical representations of different vegetation and salinity scenarios across the whole basin can be generated by using large GIS based models. These images can be a powerful means for exploring the water quality outcomes of various land use options.

In another exercise, based on IQQM, Herron et al. (2002) explored the impact of a moderate level of reafforestation in the Macquarie catchment, New South Wales, Australia. The reafforestation has been proposed as a solution to a range of environmental health issues as well as an alternative commercial enterprise for farmers along the river system. The aim was to explore the implications of the likely reduction in streamflow following the planting, to facilitate policy development and government planning on the issue. Because of the size of the catchment and nature of the problem, Herron et al. had to build an assembly of models to undertake the exercise. A lumped parameter catchment model (the "Sacramento Model", Burnash et al., 1984), was coupled to IQQM, a flow routing model (DLWC, 1995). Tree planting scenarios were developed by using a simulated forest growth surface produced by a 1-dimensional forest stand model (3-PG, Landsberg and

Waring, 1997) across the domain, with several future climate scenarios generated by a climate model for the Australian environment (OzClim) (CSIRO, 1996; Page and Jones, 2001). An empirical model (Zhang et al., 2001) was used to predict the outflow from each subcatchment according to its proportion of forest cover. Thus a composite of hydrological, forest growth, and climate models was used, taking input from three other models; one giving the forcing climate data, and two effectively giving the parameter sets for the hydrological data, at least one of which is unsuited to the role. The authors found that, at an annual time scale, the hydrological model combination gave little difference between the streamflow from grassed and treed catchments so the results were ‘‘adjusted’’.

Herron et al. (2002) found that simulations with their complex combination of models led to the conclusion that tree planting is likely to reduce the streamflow overall; a conclusion with which many would agree. The authors of the paper (Herron et al.) are all good scientists, and they are careful to state their conclusions as simulation results, not truth. Indeed, they also take some trouble to clarify that the choice of models was governed by the clients' familiarity which ‘‘increased the acceptability of the results’’! (my exclamation). It would be better that the reputation and expertise of the scientists guide the choice, but then that is, perhaps, the world we work in. The paper is an example of a very complicated application of a collection of ‘‘rules of thumb’’. The only data used were the precipitation and streamflow used to calibrate the hydrological model combination, and their projections went beyond the calibrated conditions with synthetic parameter distributions (albeit reasonable ones). The result was a report that probably precipitated some sombre discussions on the fate of the river dependent ecosystems along the course of the river. Presenting the results of this exercise to policy makers may well have raised awareness, focused concern, and perhaps modified policy and resource allocation. This would be a good outcome.

Another example of the use of models as a substitute for data was the application of a sophisticated biophysical catchment model to simulate the growth of agroforestry plots in different arrangements on different hillslopes, with different soils, salinity and climates (Silberstein et al., 2001). The model TOPOG (Vertessy et al., 1993, 1996), couples catchment hydrology to vegetation growth which depends on soil, water, salt and climatic conditions. Silberstein et al. (2002) developed design guidelines for agroforestry on hillslopes in Australia partly based on analysis of 9000 simulations of trials (Silberstein et al., 2001) to distil out the overriding principles. There is no way resources would support such a set of trials in a real catchment, and decision makers would not wait the 10–20 years required to make the measurements and study the outcomes. These simulations relied on a sophisticated biophysical hydrological model and intensive analysis to extract the emergent properties from the results. This proved a huge task, but nothing compared to the job of carrying out real experiments. For this study, data were available from only two experiments that could be used to replicate two of the 9000 simulations. These, and other field measurement exercises not directly related, were used to constrain the model, but clearly many of the simulations would have been outside observed bounds, even if the parameters defining the simulations themselves were not. Indeed, the simulations run produced very interesting results, but what was lacking was adequate data to determine whether some of the results would really occur. Fig. 2 shows the results of a set of simulations in which tree belts were simulated to be various widths and centred at different places on the hillside. The results shown are the wood growth at different places within the belts, and what is immediately apparent is the degree of variability between the simulations and at different places within belts in the same simulations. The results are discussed by Silberstein et al. (2001), and can be reasonably explained, however, the response of trees in these simulations is dependent on the combination of soil, climate (especially periodicity of storm events and seasonal distribution of rainfall), and slope, and, significantly, the representation of root activity and distribution within the soil. In most cases we have little way of knowing how real these results are. The authors presumed that the results were reasonable, but a significant input of real data is required to test many of the simulations for reality. There are a few experiments under way but it will take quite a while for results to be useful (Dr Kelvin Montague, NSW State Forests, and Dr Richard Harper, Forest Products Commission, WA, personal communications).

These examples have been chosen simply to illustrate the point that scenario modelling is undertaken with good intentions, but data are needed to constrain the discussion and improve the understanding and limitations of the results. In the case of simulations under future climate scenarios, these data are not, and will not, be available within the timeframe of the model's use, and errors in prediction are unlikely to be detected until long after the modellers have moved on from current positions. I suggest that applying the results in other than a precautionary planning manner is dangerous.

6. Models need data

As models gain complexity, or expand the processes represented, the demand for data to calibrate and validate them increases. At the same time, as our technology improves and we have the ability to measure more attributes with greater precision, models expand to make

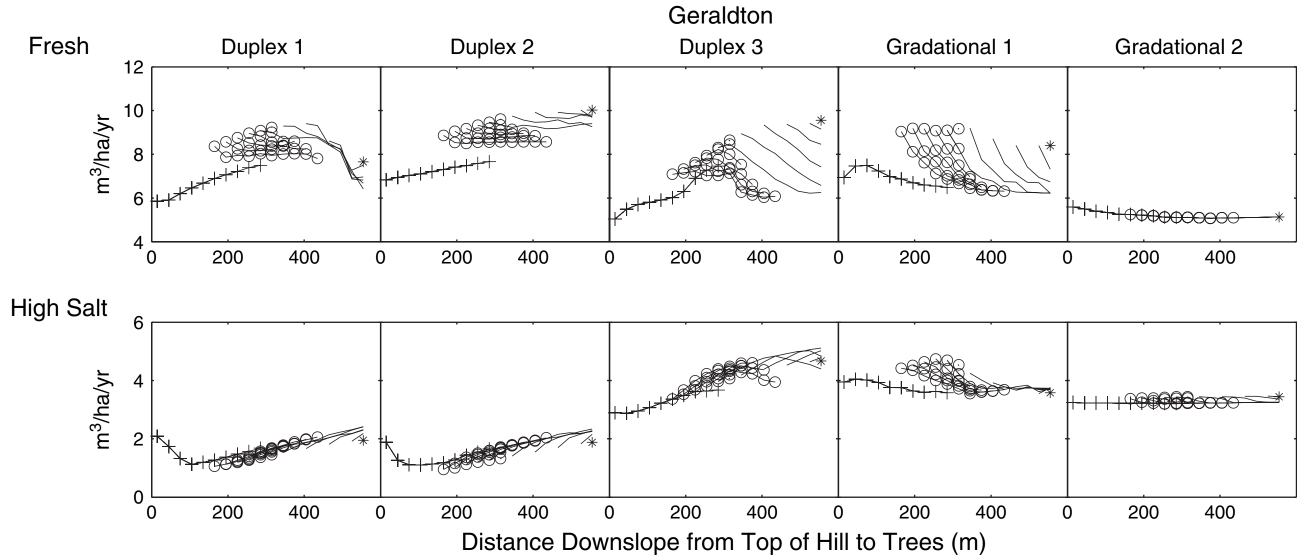

Fig. 2. Total growth of agroforestry plantations under the climatic conditions in Geraldton, Western Australia. Results from 150 simulations, with each point representing growth of trees at that location on the hillslope are shown. Each panel is for a different soil profile, with the top row showing soil with no salt and the bottom row depicting high levels of soil salinity. The lines of common type indicate the width and spread of the tree belts for that simulation. The horizontal axis shows position on the hillside, with 0 being at the top and 600 being at the bottom of the hill (units are m).

use of these opportunities. For example, the host of models being driven by remotely sensed data, especially surface temperature and vegetation cover data. Data availability now is truly impressive. In southern Western Australia a true 2 m digital elevation model (DEM) is available to generate catchment networks, and is being used to develop a catchment flow network for the 120,000 km^2 Avon Drainage basin (Riasat Ali, personal communication). Over the whole of Australia the SILO data resource (Bureau of Meteorology and Queensland Department of Natural Resources and Mining http://www1.ho.bom.gov.au/silo/) has over 100 years of daily rainfall and 45 years of daily climate data from thousands of stations across the country, and anywhere in between. However, these examples are of data used to drive hydrological models, not to test them.

The scientific literature is full of modelling and data studies, and while there are many studies exploring data without models, there are far fewer exploring models without data – at least in the natural sciences. Studies comparing models to data (Moore and Mein, 1976; Wheater et al., 1993; Grayson and Chiew, 1994; Ye et al., 1997; and so on) seek to assess model performance by comparing model output to data. While there is no known "Universal Law of Hydrology" to guide the development of models (McCuen, 1997; Beven, 2001), surely the aim of our science is to find such a model that will work in any catchment, with any climate, and any vegetation; this requires continual observation and analysis to develop and refine our understanding (our model).

The hydrological community has divided into "modellers" and "experimentalists", and despite the obvious benefits to both groups, calls to combine forces (Dunne, 1983; Klemes, 1986) are rarely observed. Data-model comparisons invariably result in differences, and these differences are the source of insight into how natural systems work. Seibert and McDonnell (2002) explore a mechanism for getting the understanding of the experimentalist into the mathematics of the modeller, through the use of "soft data", or experimentalists' understanding of "how the catchment works", in addition to the "hard data" of streamflow, climate and soil parameters modellers are used to. They conclude that it is better to be "less right for the right reason" than "right for the wrong reason".

6.1. Modelling difficult environments

Much of Australia is arid or semi-arid, conditions typically much harder to model than humid temperate catchments like those in most of Europe and North America. The hydrology of these regions is the subject of increasing interest (Pilgrim et al., 1988; Karnieli and Ben-Asher, 1993; Smakhtin, 2001), particularly as much of the world's population lives in such regions. Many of these catchments (and most in Australia) also have low relief, with for example, the average slope in the Murray-Darling Basin being around 1:10,000 (and the Basin having an area 10^6 km^2). These catchments typically have rainfall to potential evaporation ratios of much less than 1 and under natural vegetation have average annual streamflow less than 10% of rainfall. Such flow statistics are very difficult to get, because the flow may be zero much of the time. Catchments in south-west

Western Australia, at 1200 mm rainfall typically have streamflow of about 10% rainfall, and at 700 mm have streamflow of 2% rainfall (for example, Ruprecht and Schofield, 1989; Silberstein et al., 1999a). Such catchments are difficult to model because the primary drivers of streamflow are the residual of evaporation losses and changes in soil moisture storage, and soil characteristics that control moisture redistribution. With fewer flow statistics, because of fewer days of flow, parameter estimation is much more difficult. Because of the high evaporation, the catchment spends a large proportion of the time with soil moisture either too low for streamflow generation or distributed such that discharge does not occur. Either way, a zero hydrograph gives no information about the water storage or the redistributions that may be taking place. With evaporation being 90% or more of rainfall, a 10% error in its estimation leads to 100% error in streamflow.

6.2. Model validation

The traditional scientific method, that is to propose a hypothesis and then set about testing whether it is untrue, seems to have been lost in many modelling studies. Klemes (2000) has discussed this in various ways over many years. In scientific tradition, a scientist takes a model, or develops one, and sets about systematically to try to *invalidate it* – the "null hypothesis". However, papers are usually written from the viewpoint of validation of a model, when in fact this can never happen. Models can be validated in an engineering sense, that within accepted boundaries they can be demonstrated to reproduce observations, and are therefore useful, but scientifically, like theories, they can never be truly validated, only invalidated.

A scientist studies how the world works. A model encapsulates some of that understanding and, for this author, the real interest comes in understanding where and why a model fails. While most modellers are concerned about validating their models there is the philosophical question of whether models can be validated or rather "verified" in the sense that they represent truth. Oreskes, et al. (1994) concluded that they cannot be verified because there are too many inherently unknowable parameters and processes. They argue that a model may be *valid*, in the sense that legal arguments are valid, in containing no logical or demonstrated flaw, and in the same sense legal arguments are valid until some time they fail a test. Konikow and Bredehoeft (1992) argued that validation is used erroneously in at least two senses: firstly, that validation = verification, and secondly that validation = truth (as the representation of reality). The US Department of Energy defines model *validation* as "*determination that the model indeed reflects the behaviours of the real world*". Thus, *validation* really only implies that the empirical observations match model results; the more observations the more confidence that they represent "truth", but that confidence is not certainty. In the same way that physical theories cannot be proven, but only disproved, models can be "supported" by, but not verified, by observations.

For example, Silberstein et al. (1999c) studied waterlogging and its relation to topography and pasture in a small catchment. They attempted to model the occurrence and used reasonably comprehensive measurements to test their model (TOPOG, described earlier). In this study, evaporation flux measurements were available and matched the model simulations well (Fig. 3), as did soil moisture storage (Fig. 4). However, the aim of the study was to quantify waterlogging and deep drainage, which were much harder to simulate (Fig. 5). The conclusion drawn from this work was that the model, that had been tested extensively in similar work, was incapable of producing the full complexity of the field processes, because it did not include a very important one, namely the changing hydraulic properties of soil through a seasonal cycle. In that study with only part of the dataset, a model calibration could have reproduced the known results, but the unmeasured ones would have remained unknown. This would put those results and any analysis and conclusions drawn from them in the world of computer games. The authors could only speculate on the relative importance of various soil changes taking place, but the model, in failing, was useful in helping ascertain where measurements were incomplete, and the magnitude of the processes that were unmeasured.

7. The Australian need for data

In Australia, as in the rest of the world, the hydrological community is beholden to the data collectors. Models are our attempt to encapsulate our understanding of the real world. They are not the real world, and without data they are simply imagination and computer games. Hydrological modelling, and any other modelling, is a much simpler, and less interesting, version of the real water world.

The need to understand our water systems was recognised by the Founding Fathers of Australia:

"On entering our duties we found … that information available regarding our rivers was meagre and fragmentary, and that in some important points public opinion was in danger of being misled by statements and theories which there was ample evidence to refute.

… we beg to recommend that the maintenance of river gauge records as extended by us should be made still more complete, and the records kept continuously and in a careful and systematic manner."

(Parliament of the Colony of New South Wales, 1887; taken from NLWRA, 2000).

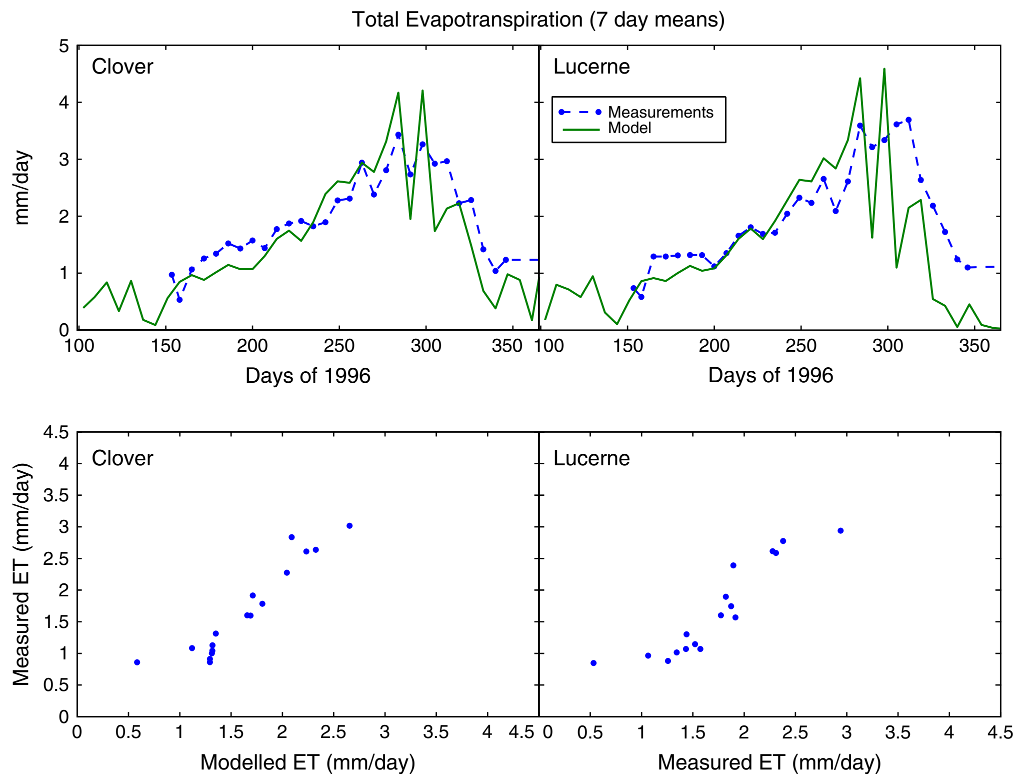

Fig. 3. Total evaporation measured using Bowen ratio instruments and modelled with TOPOG at a pasture trial in south-west Western Australia. Taken from Silberstein et al. (1999c).

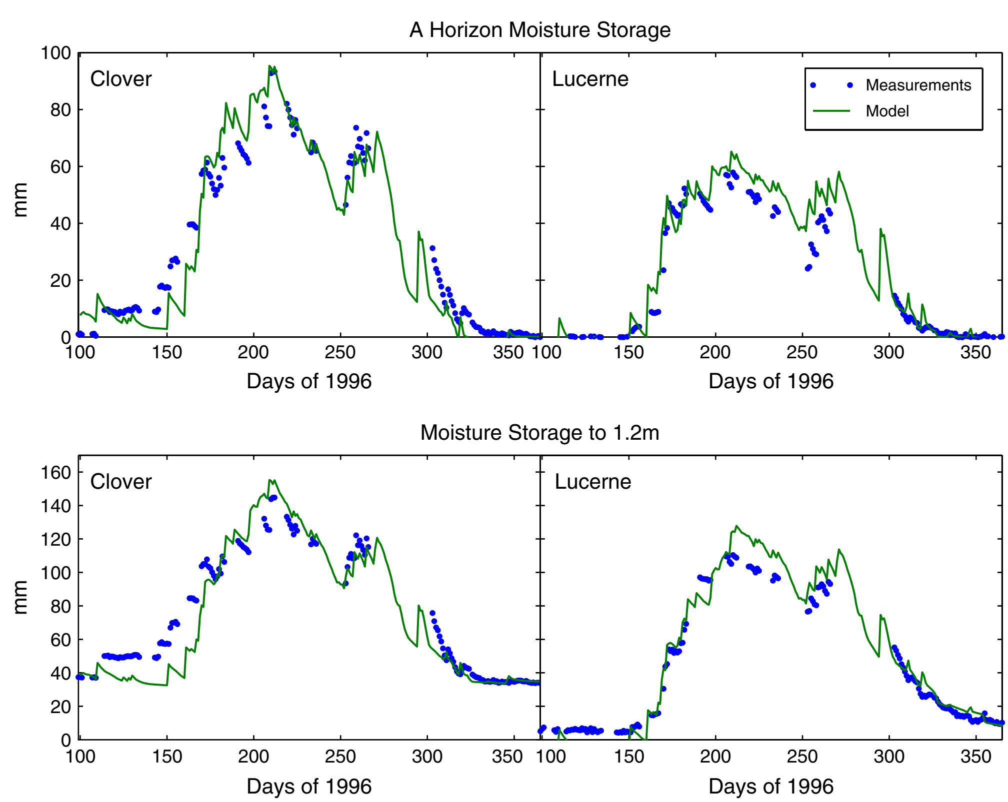

Fig. 4. Soil moisture storage in the A soil horizon and to 1.2 m depth measured using neutron moisture meters and time domain reflectometry, and modelled with TOPOG at a pasture trial in south-west Western Australia. Taken from Silberstein et al. (1999c).

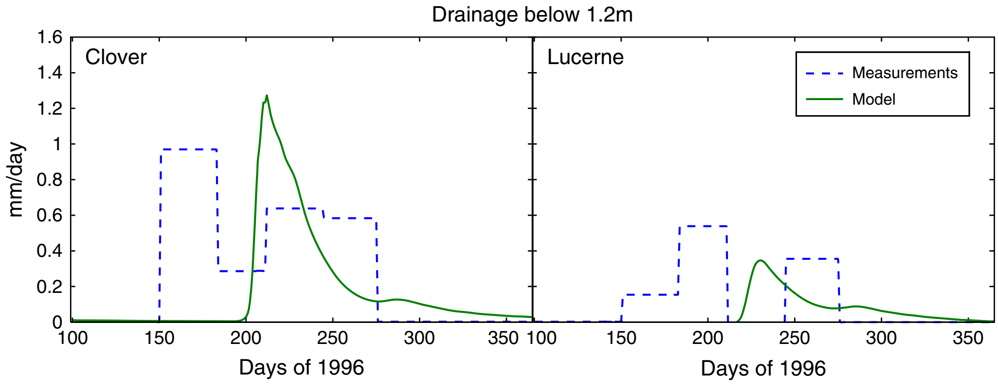

Fig. 5. Monthly deep drainage below 1.2 m depth in the soil at a pasture trial in south-west Western Australia. The "measurements" are actually calculations from soil water balance, the model is TOPOG, taken from Silberstein et al. (1999c).

It is my contention that our current need to understand our hydrological systems through measurement is just as urgent as it was then. In many parts of the world, not the least Australia, hydrological data collection is being reduced. In 2002 enquiries of officers of State hydrological agencies in almost every Australian State reported to me that data collection was being reduced. The attitude of those in positions to make these decisions seems to be that we have enough data, and that not enough use is made of it. The view is that we do not need to collect data because "we can model" and, in any case, "the main use of data is to calibrate models". These statements were made to the author by senior people charged with collecting hydrological data in Australia, and both of them indicate a significant lack of understanding of how science works, and how we should be using it to improve the management of our environment. Data are the real world. We do not need data if we do not care about the real outcome – if we are happy with computer games.

7.1. Data are needed in a changing environment

Modelling the impacts of land clearing on stream changes, flows and salinity would be impossible without data. The change to flow regimes following clearing of native forest for agriculture, such as those in the Collie research catchments, in the south-west of Western Australia (Mauger et al., 2001; Silberstein et al., 2003a, 2005), that have converted ephemeral streams to perennial streams are completely beyond the data boundaries prior to these impacts. No model can demand great confidence when it goes beyond the boundaries of its data and calibration.

Environmental data are the raw materials for environmental audits. Much like financial audits, continuous collection would allow a chance of redressing the problem if the balance sheet goes awry. The Australian Government has, on behalf of Australian taxpayers, embarked on a $1400M experiment, that is the National Action Plan on Salinity and Water Quality (http://www.napswq.gov.au). This is the second such experiment after the Natural Heritage Trust (http://www.nht.gov.au/). It is imperative that data are collected to confirm whether the new experiment is a success, when it is argued that the previous one had minimal environmental outcomes (http://www.acfonline.org.au), and there was virtually no monitoring to confirm whether this was the case or not. Large corporations undertaking financial outlays of this magnitude require checks and balances to audit the process and to stage the process to ensure that it is managed wisely. It is only prudent to learn as the process is undertaken, and there should be mechanisms in place to improve outcomes as the programme proceeds. Australia currently spends only a few per cent of this amount on water monitoring. I suggest that a small additional investment would ensure that we could account for how well the rest was spent. We can use models to explore the possibilities but it is only data that will tell us which possibility has become a reality.

The International Association of Hydrological Sciences has declared 2003–2012 a Decade for Prediction in Ungauged Basins (PUB) (http://cee.uiuc.edu/research/pub/default.asp). These are overdue activities and it is to be hoped that the result is a substantial raising of the profile of the science and the benefits to humanity. In a sense the aim of the PUB initiative is to develop hydrological models that can work in catchments with minimal data, but the clearly stated modus operandi in the Science Plan (IAHS, 2003) is to maximise data collection and analysis. There is a focus on developing new observational tools that will enlighten us on the inner workings of catchments; that we may better understand and predict their responses to forcing, and thereby better manage for environmental and human outcomes. The PUB initiative has included clear community service goals in its aims, and emphasises the role that data will play in developing hydrological understanding for the betterment of societies across the globe.

8. Concluding remarks

Models are enormously useful as test beds for ideas and for exploring the implications of our understanding of natural systems. They are extremely valuable as data processing and analysis aids, often showing up data errors and inconsistencies that might otherwise have gone unnoticed. Models are also useful for exploring scenarios that cannot be tested in the real world. However, while this last use is a rapidly expanding one, it is also the most dangerous, as high level managers appreciate the nice graphics and, possibly, simplistic sets of options, it can be easy to lose sight of the limitations of the process that generated them. It is in this mode that models are often run outside their tested bounds, and by definition little or no data are available to constrain the scenario results. If we are to continue to learn about, and improve our management of, our environment, we must continue to observe it – that means collect data. Modelling is an important accompaniment to measurement, but is no substitute for it; science requires observation, and without that we will cease to progress in understanding our environment, and therefore in managing it appropriately.

Acknowledgments

The genesis of this paper, and the comments received from senior public servants that have prompted it, occurred while the author was undertaking the CSIRO Project Leaders' Programme. The author wishes to thank Giuseppe Gigliotti for his encouragement and guidance during that time. The author also wishes to thank Bernadette Waugh for cheerfully and enthusiastically helping track down references on this topic. Dr Stephen Charles and Mr Warrick Dawes made helpful comments on the paper, as did two anonymous referees.

References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E., Rasmussen, J., 1986. 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.

Beven, K.J., 1997. Distributed Hydrological Modelling: Applications of the TOPMODEL concept. John Wiley and Sons, Chichester, UK.

Beven, K.J., 2000. Uniqueness of place and non-uniqueness of models in assessing predictive uncertainty. In: Bentley, et al. (Eds.), Computational Methods in Water Resources XIII. Calgary, Alberta, Canada, pp. 1085–1091.

Beven, 2001. Rainfall-Runoff Modelling: the Primer. John Wiley and Sons, Chichester, UK.

Beven, K.J., Calver, A., Morris, E.M., 1987. The Institute of Hydrology Distributed Model. Institute of Hydrology, Wallingford, England.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences 24 (1), 43–69.

Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R., Freer, J., 1995. TOPMODEL, In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Colorado, USA, pp. 627–668.

Boughton, W.C., 1995. An Australian water balance model for semiarid watersheds, In: Water Research and Management in Semiarid Environment, Proceedings of an international symposium, Tucson, Arizona, USA. 1–3 November, 1994.

Brakensiek, D.L., 1967. Kinematic flood routing. Transactions of the American Society of Agricultural Engineers.

Burnash, R.J.E., Ferral, R.L., McGuire, R.A., 1984. A Generalised Streamflow Simulation System. Joint Federal-State River Forecast Centre, Sacramento, California, USA.

CSIRO, 1996. OzClim: A Climate Scenario Generator and Impacts Package for Australia. Commonwealth Scientific and Industrial Research Organisation, Atmospheric Research, Melbourne, Australia. <http://www.dar.csiro.au/res/cm/ozclim.htm>.

Darcy, H., 1856. Les Fontaines Publique de la Ville de Dijon, Paris, Dalmont.

Dawes, W.R., Zhang, L., Dyce, P.A., 1998. WAVES V3.5 User Manual. CSIRO Land and Water, Canberra, ACT, 2601.

DLWC, 1995. Integrated Quantity-Quality Model (IQQM) Reference Manual. Department of Land and Water Conservation, Sydney, NSW, Australia.

Dooge, J.C.I., 1988. Hydrology in perspective. Hydrological Sciences Journal 33 (1), 61–85.

Dunne, T., 1983. Relation of field studies and modeling in the prediction of storm runoff. Journal of Hydrology 65, 25–48.

Farmer, D., Sivapalan, M., Jothityangkoon, C., 2003. Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: downward approach to water balance analysis. Water Resources Research 39 (2), 1035. doi:10.1029/2001WR000328.

Fowkes, N.D., Mahony, J.L., 1994. An Introduction to Mathematical Modelling. John Wiley and Sons, Chichester, UK.

Gash, J.H.C., Wright, I.R., Lloyd, C.R., 1980. Comparative estimates of interception loss from three coniferous forests in Great Britain. Journal of Hydrology 48, 89–105.

Grayson, R.B., Chiew, F.H.S., 1994. An approach to model selection. In: Water Down Under 94 Symposium. Institution of Engineers, Australia, pp. 507–512.

Grayson, R.B., Moore, I.D., McMahon, T.A., 1992. Physically based hydrologic modeling. 2. Is the concept realistic? Water Resources Research 28, 2659–2666.

Henderson-Sellers, A., Yang, Z.L., Dickinson, R.E., 1993. The project for intercomparison of land-surface parameterization schemes. Bulletin of the American Meteorological Society 74 (7), 1335–1349.

Herron, N., Davis, R., Jones, R., 2002. The effects of large-scale afforestation and climate change on water allocation in the Macquarie River catchment, NSW, Australia. Journal of Environmental Management 65, 369–381.

Horton, R.E., 1933. The role of infiltration in the hydrologic cycle. Transactions of the American Geophysical Union 14, 446–460.

IAHS, 2003. International Association of Hydrological Sciences (Web Page). Available at: <http://www.cig.ensmp.fr/~iahs/>, (accessed 01.04.05.).

Karnieli, A., Ben-Asher, J., 1993. A daily runoff simulation in semi-arid watersheds based on soil water deficit calculations. Journal of Hydrology 149, 9–25.

Klemes, V., 1986. Dilettantism in hydrology; transition or destiny? Water Resources Research 22 (9), 177S–188S.

Klemes, V., 2000. Common Sense and Other Heresies. Canadian Water Resources Association, Ontario, Canada.

Konikow, L.F., Bredehoeft, J.D., 1992. Groundwater models cannot be validated. Advances in Water Resources 15 (1), 75–83.

Landsberg, J.J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology and Management, 95, 209–228.

The Macquarie Dictionary, second revision. The Macquarie Library Pty Ltd ISBN: 0 949757 41 1.

Mauger, G.W., Bari, M., Boniecka, L., Dixon, R.N.M., Dogramaci, S.S., Platt, J., 2001. Salinity Situation Statement – Collie River. No. WRT 29. Water and Rivers Commission, Perth, Western Australia.

McCuen, R.H., 1997. Hydrologic Analysis and Design, second ed. Prentice Hall, pp. 814.

McVicar, T.R., Jupp, D.L.B., 1999. Estimating one-time-of-day meteorological data from standard daily data as inputs to thermal remote sensing based energy balance models. Agricultural and Forest Meteorology 96, 219–238.

Monteith, J.L., 1965. Evaporation and environment. In: Fogg, G.E. (Ed.), The State and Movement of Water in Living Organisms, Symposium of the Society of Experimental Biology. Academic Press, New York, pp. 205–234.

Moore, I.D., Mein, R.G., 1976. Evaluating rainfall/run-off model performance. Journal of Hydraulics Division, ASCE, 102, 1390–1395.

NLWRA, 2000. National Land and Water Resources Audit (Web Page), Available at: <http://www.nlwra.gov.au/>, 2000 (accessed 1.04.05.).

Oreskes, N., Shrader-Frechette, K., Kenneth, B., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263, 641–646.

Page, C.M., Jones, R.N., 2001. OzClim: the development of a climate scenario generator for Australia. In: Ghassemi, F., Whetton, H., Little, R., Littleboy, M. (Eds.), MODSIM 2001: International Congress on Modelling and Simulation: Proceedings, Canberra, ACT, pp. 667–671.

Passioura, J.B., 1996. Simulation models: science, snake oil, education, or engineering? Agronomy Journal 88, 690–694.

Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. Proceedings of Royal Society A 193, 120–145.

Pilgrim, D.H., Chapman, T.G., Doran, D.G., 1988. Problems of rainfall-runoff modelling in arid and semiarid regions. Hydrological Sciences Journal 33, 379–400.

Pitman, A.J., Henderson-Sellers, A., 1998. Recent progress and results from the project for the intercomparison of land surface parameterization schemes. Journal of Hydrology 212/213 (1/4).

Reynolds, J.F., Acock, B., 1985. Predicting the response of plants to increasing carbon dioxide: a critique of plant growth models. Ecological Modelling 29, 107–129.

Richards, L.A., 1931. Capillary conduction of liquids in porous mediums. Physics 1, 318–331.

Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. Water Resources Research 8, 1204–1213.

Ritchie, J.T., Otter-Nacke, S., 1985. Testing and Validating the CERES-Wheat Model in Diverse Environments. AgRISTART Publ. No YM-15-00407. NTIS, Springfield, VA, USA.

Ruprecht, J.K., Schofield, N.J., 1989. Analysis of streamflow generation following deforestation in southwest Western Australia. Journal of Hydrology 105, 1–17.

Rutter, A.J., Kershaw, K.A., Robins, P.C., Morton, A.J., 1971. A predictive model of rainfall interception in forests. I: derivation of the model from observations in a plantation of Corsican Pine. Agricultural Meteorology 9, 367–384.

Saint-Venant, Barre de, 1871. Theory of unsteady water flow, with application to river floods and to propagation of tides in river channels. French Academy of Science 73 (148–154), 237–240.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeller in catchment hydrology: use of soft data for multicriteria model calibration. Water Resources Research 38 (11), 2301–2314.

Silberstein, R.P., Sivapalan, M., Wyllie, A., 1999a. On the validation of a coupled water and energy balance model at small catchment scales. Journal of Hydrology 220, 149–168.

Silberstein, R.P., Vertessy, R.A., Morris, J.D., Feikema, P.M., 1999b. Modelling the effects of soil moisture and solute conditions on long-term tree growth and water use: a case study from the Shepparton irrigation area, Australia. Agricultural Water Management 39 (2–3), 283–315.

Silberstein, R.P., Hatton, T.J., Ward, P., Williamson, D.R., Bartle, G., Lambert, P., Dunin, F.X., Micin, S.F., Mungai, D., Ward, B., 1999c. Modelling drainage and transient waterlogging in an agricultural catchment, In: Water 99, Proceedings of the 25th Hydrology and Water Resources Symposium and Second International Conference on Water Resources and Environmental Research, Brisbane, pp. 999–1004.

Silberstein, R.P., Vertessy, R.A., McJannet, D., 2001. A parameter space and simulation response surface for agroforestry design. In: MODSIM, 2001. International Congress on Modelling and Simulation, Canberra, December, 2001. The Modelling and Simulation Society of Australia and New Zealand Inc, pp. 1889–1894.

Silberstein, R.P., Vertessy, R.A., McJannet, D., Hatton, T.J., 2002. Tree belts on hillslopes. In: Stirzaker, R.J., Vertessy, R.A., Sarre, A. (Eds.), Trees, Water and Salt – An Australian Guide to Using Trees for Productive Farms and Healthy Catchments. Rural Industries Research and Development Corporation, Canberra, ACT.

Silberstein, R.P., Adhitya, A., Dabrowski, C., 2003a. Changes in Flood Flows, Saturated Area and Salinity Associated with Forest Clearing for Agriculture. Technical Report 03/1. CRC for Catchment Hydrology, Melbourne, Australia.

Silberstein, R.P., Sivapalan, M., Viney, N.R., Held, A., Hatton, T.J., 2003b. Modelling the energy balance of a natural jarrah (*Eucalyptus marginata*) forest in Western Australia. Agricultural and Forest Meteorology 115 (3–4), 202–232.

Silberstein, R.P., Gargett, T., Adhitya, A., Best, A., Hickel, K., 2005. The Effect of Clearing of Native Forest on Flow Regime Technical Report 04/4. CRC for Catchment Hydrology, Melbourne, Australia.

Sivapalan, M., Ruprecht, J.K., Viney, N.R., 1996a. Water and salt balance modelling to predict the effects of land-use changes in forested catchments. 1. Small catchment water balance model. Hydrological Processes 10, 393–411.

Sivapalan, M., Viney, N.R., Jeevaraj, C.G., 1996b. Water and salt balance modelling to predict the effects of land-use changes in forested catchments. 3. The large catchment model. Hydrological Processes 10, 429–446.

Smakhtin, V.Y., 2001. Low flow hydrology: a review. Journal of Hydrology 240, 147–186.

Vertessy, R.A., Hatton, T.J., O'Shaughnessy, P.J., Jayasuriya, M.D.A., 1993. Predicting water yield from a mountain ash forest catchment using a terrain analysis based catchment model. Journal of Hydrology 150, 665–700.

Vertessy, R.A., Dawes, W.R., Zhang, L., Hatton, T.J., Walker, J., 1996. Catchment-scale hydrologic modelling to assess the water and salt balance behaviour of eucalypt plantations, Technical Memorandum No. 96/2, CSIRO Div. Water Resources.

Viney, N.R., Bates, B.C., Charles, S.P., Webster, I.T., Bormans, M., Aryal, S.K., 2003. Impacts of climate variability on riverine algal blooms. MODSIM 2003, International Congress on Modelling and Simulation, Townsville, Queensland, July 2003.

Watson, F.G.R., Vertessy, R.A., Grayson, R.B., 1999. Large-scale modelling of forest hydrological processes and their long-term effect on water yield. Hydrological Processes 13 (5), 689–700.

Wheater, H.S., Jakeman, A.J., Beven, K.J., Beck, M.B., McAleer, M.J., 1993. Progress and Directions in Rainfall-Runoff Modelling, Modelling Change in Environmental Systems. John Wiley and Sons, Chichester, UK.

Wu, H., Rykiel Jr., E.J., Hatton, T.J., Walker, J., 1994. An integrated rate methodology (IRM) for multi-factor [plant] growth rate modelling. Ecological Modelling 73, 97–116.

Ye, W., Bates, B.C., Viney, N.R., Sivapalan, M., Jakeman, A.J., 1997. Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. Water Resources Research 33, 153–166.

Zhang, L., Dawes, W.R., Walker, G.R., 2001. The response of mean annual evapotranspiration to vegetation changes at catchment scale. Water Resources Research 37, 701–708.