HYDROLOGICAL PROCESSES
Hydrol. Process. **20**, 1057–1073 (2006)
Published online 18 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.5933

Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability

Wenzhi Cao,[1,2] William B. Bowden,[3] Tim Davie[1]* and Andrew Fenemor[4]

[1] *Landcare Research, PO Box 69, Lincoln 8152, New Zealand*
[2] *Environmental Science Research Centre, Xiamen University, Xiamen, Fujian 361005, People's Republic of China*
[3] *School of Natural Resources, University of Vermont, Burlington, VT 05405, USA*
[4] *Landcare Research, Private Bag 6, Nelson 7001, New Zealand*

Abstract:

Many methods developed for calibration and validation of physically based distributed hydrological models are time consuming and computationally intensive. Only a small set of input parameters can be optimized, and the optimization often results in unrealistic values. In this study we adopted a multi-variable and multi-site approach to calibration and validation of the Soil Water Assessment Tool (SWAT) model for the Motueka catchment, making use of extensive field measurements. Not only were a number of hydrological processes (model components) in a catchment evaluated, but also a number of subcatchments were used in the calibration. The internal variables used were PET, annual water yield, daily streamflow, baseflow, and soil moisture. The study was conducted using an 11-year historical flow record (1990–2000); 1990–94 was used for calibration and 1995–2000 for validation. SWAT generally predicted well the PET, water yield and daily streamflow. The predicted daily streamflow matched the observed values, with a Nash–Sutcliffe coefficient of 0·78 during calibration and 0·72 during validation. However, values for subcatchments ranged from 0·31 to 0·67 during calibration, and 0·36 to 0·52 during validation. The predicted soil moisture remained wet compared with the measurement. About 50% of the extra soil water storage predicted by the model can be ascribed to overprediction of precipitation; the remaining 50% discrepancy was likely to be a result of poor representation of soil properties. Hydrological compensations in the modelling results are derived from water balances in the various pathways and storage (evaporation, streamflow, surface runoff, soil moisture and groundwater) and the contributions to streamflow from different geographic areas (hill slopes, variable source areas, sub-basins, and subcatchments). The use of an integrated multi-variable and multi-site method improved the model calibration and validation and highlighted the areas and hydrological processes requiring greater calibration effort. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS physically based distributed hydrological models; calibration and validation; soil and water assessment tool; spatial variability

INTRODUCTION

Physically based, distributed hydrological models (PDHMs), whose input parameters have a physical interpretation and explicit representation of spatial variability (Abbott *et al*., 1986), are increasingly being used to solve complex problems in water resources applications (Sorooshian and Gupta, 1995), including environmental impacts of land-use changes, effects of climate change on water resources, and water planning and management in a catchment. However, problems with PDHMs include a lack of sufficient data to characterize spatial variability fully, scale problems of integration of field measurements and model parameter element, and imperfect representations of real processes in models (Beven, 1989, 1993, 2002; Grayson *et al*., 1992a). These factors invariably result in a requirement for model calibration and validation (Refsgaard, 1997; Anderton *et al*., 2002).

* Correspondence to: Tim Davie, Landcare Research, PO Box 69, Lincoln 8152, New Zealand. E-mail: daviet@landcareresearch.co.nz

Received 1 October 2003
Accepted 9 February 2005

In many cases, the appropriate values for a model parameter are determined through a trial-and-error process in which a small number of key parameters are manipulated in an attempt to achieve the desired response (Gupta *et al*., 1998; Anderton *et al*., 2002). The model calibration is usually based on a comparison between the simulated and observed streamflow, primarily depending on a modeller's hydrological expertise. Nevertheless, the potential for equifinality or non-uniqueness in complex, spatially distributed models with numerous calibration parameters has shown that a large number of alternative parameterizations can produce acceptable results. This is particularly true when a single variable, e.g. outlet streamflow, is selected as the sole calibration criterion (e.g. Beven, 1993, 1996, 2001).

The single-criterion method has been found to be limited when calibrating a complex numerical model with many parameters (Gupta *et al*., 1999; Anderton *et al*., 2002). The equifinality problem is of particular importance in PDHMs owing to their distributed structure and the huge number of parameter values often required to be estimated and optimized. To tackle this problem, different calibration methods for PDHMs have been developed, e.g. the generalized likelihood uncertainty estimation (Beven and Binley, 1992). Automatic calibration procedures with an optimization strategy include the genetic algorithms (Wang, 1991) and the shuffled complex evolution (SCE-UA) global optimization algorithm (Duan *et al*., 1992). However, these methods are time consuming and computationally intensive; thus, only a small set of input parameters can be optimized and the methods can optimize to unrealistic parameter values unless appropriate, physically realistic constraints are included in the algorithm.

A subsystem approach to calibrating internal state variables, such as evapotranspiration and baseflow, can be integrated into a model calibration and validation process. This multi-variable calibration method can fully use the field measurements, and it has been suggested as an effective methodology for reducing uncertainty in parameter identification for PDHMs (Fenemor, 1988; Grayson *et al*., 1992b; Anderton *et al*., 2002; Bergstrom *et al*., 2002), particularly in a large catchment with high heterogeneity and spatial variability.

In this study, the physically based, distributed model Soil Water Assessment Tool (SWAT; Arnold *et al*., 1998) has been applied at a large scale in the Motueka River catchment in New Zealand. A multi-variable and multi-site approach to calibration and validation of SWAT has been used through trial-and-error processes. Not only have internal hydrological processes in the model been evaluated, but also a number of subcatchments have been used in the calibration. This paper presents the use of multiple variables and multiple sites for calibration and validation of a model within a large, complex catchment; it is not aimed at the estimation of parameter uncertainty. Fohrer *et al*. (2001) have investigated parameter uncertainty in SWAT through sensitivity analysis in a small artificial catchment. The results showed that model sensitivity to land-cover-related parameters varied with both time and model output variables.

The hydrological components used for calibration and validation in this study were precipitation, temperature, potential evapotranspiration (PET), total water yield, and baseflow. In addition, performance of the model in six subcatchments was used to calibrate and validate the model. Extensive field measurements have been used to calibrate and internally validate the SWAT model in the Motueka catchment.

CATCHMENT DESCRIPTION

The Motueka River basin is situated at the north of the South Island of New Zealand. The river drains an area of 2075 km^2 and has a main stem length of approximately 36 km. It provides 65% of the major freshwater flow into Tasman Bay. Altitude exceeds 1600 m in the upper catchments of the two major tributaries, the Motueka and Wangapeka. Two-thirds of the catchment is steep country, with slopes exceeding 27%. The catchment has a complex mixture of geology and land use, and water availability is a critical issue with competition among multiple in-stream and off-stream users.

Land uses in the Motueka catchment comprise exotic forestry, mainly *Pinus radiata* (covering 25% of the catchment area), sheep and beef farming (19%), and limited but increasing dairying. Horticulture (mainly pip fruit, berry fruit, hops, vegetables) occupies a small, but expanding, area. A large area of the catchment

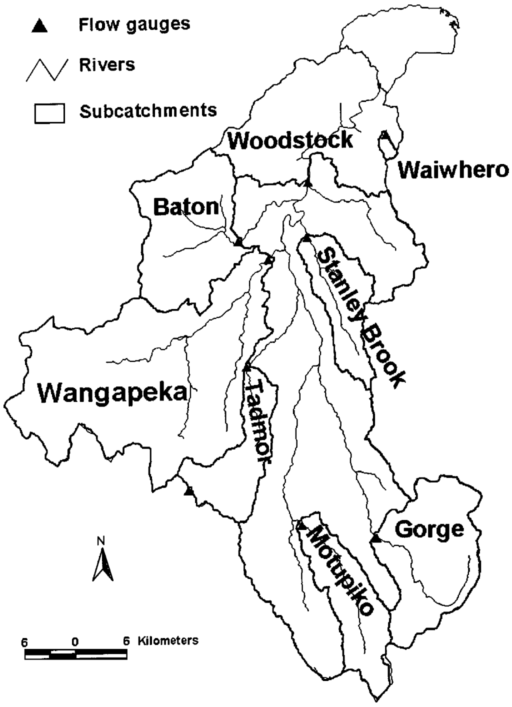

Figure 1. Subcatchments in the Motueka catchment

is conservation estate with indigenous forest, scrub and tussock grassland (55%). This is mainly in the high-rainfall headwaters of the western tributaries and upper Motueka.

Geologically, the upper Motueka headwaters are underlain by Permian age ultramafic and sedimentary rocks. The western tributaries are underlain by a complex mixture of sedimentary and igneous rocks dating from the Cambrian through to Miocene ages. The middle and lower reaches of the main stem and eastern tributaries of the Motueka are underlain by thick layers of glacial outwash gravels and younger alluvium.

A digital thematic map of land cover interpreted mainly from summer 1996–97 satellite imagery was used to define land use in the Motueka catchment, and the physical and chemical properties (Chittenden *et al.*, 1966) were derived from the National Soils Database and Land Resources Inventory (Wilde *et al.*, 1999, unpublished report).

The Motueka catchment has been divided into seven nested subcatchments based on measured flow records (Figure 1), and these have a wide variety of land uses, land covers, geology, and soil types. Six subcatchments plus the Motueka catchment at Woodstock were used to calibrate and validate the model spatially (Figure 1 and Table I).

MODEL DESCRIPTION

SWAT is a PDHM that operates on a daily time-step (Arnold *et al.*, 1998). A catchment is first split into sub-basins according to the terrain and river channels, and then into multiple hydrological response units (HRUs) based on the soil and land cover types within the sub-basins. An HRU is a fundamental spatial unit upon which SWAT simulates the water balance. A comprehensive description of all the components in SWAT can be found in the literature (e.g. Arnold and Allen, 1996; Arnold *et al.*, 1998; Srinivasan *et al.*, 1998).

Table I. Motueka catchment and subcatchments

Catchment and subcatchments (flow gauge names)	Area (km^2)	Area (%)[a]
Upper Motueka (Gorge)	163·0	9·3
Motupiko (Christies)	105·4	6·0
Tadmor (Mudstone)	88·0	5·0
Wangapeka (Walters Peak)	479·0	27·4
Stanley Brook (Barkers)	81·6	4·7
Baton (Baton Flats)	168·0	9·6
Motueka (Woodstock)[b]	1765·6	100·0

[a] Percentage of the Motueka catchment above Woodstock.
[b] The Woodstock flow gauge provides the last river measurement for the catchment, although there is still about 309 km^2 below this point.

Briefly, the hydrological processes modelled in SWAT are precipitation, surface runoff, soil and root-zone infiltration, evapotranspiration and soil and snow evaporation, and baseflow.

As part of this study, three major adjustments have been made to the standard SWAT model. These are the way curve numbers are used and the interpolation of precipitation and temperature among gauges and stations.

Daily precipitation prediction

Precipitation is the key input variable that drives flow and mass transport in hydrological systems. The spatial variability of precipitation and the accuracy of the precipitation input are critical to the use of hydrological models (e.g. Beven and Hornberger, 1982; Hamlin, 1983; Shah *et al.*, 1996). The annual precipitation in the Motueka catchment ranges from about 950 mm to more than 3500 mm, with high variability due to a complex, rugged terrain. The arrangement of the mountains and the predominately westerly airflows result in a strong precipitation gradient from west to east in the catchment. There is an irregular spacing of precipitation gauges around the Motueka catchment, with 18 gauges with long enough records for use in a long-term modelling study. Unfortunately, these are distributed mainly in the lower parts of the catchment (Figure 2). However, SWAT assigns the climate parameter values (e.g. precipitation, temperature) obtained from the closest station to a sub-basin.

A separate preprocessing model has been developed to predict the daily precipitation based on the 18 gauges and estimated annual isohyets. Assuming the precipitation at a point is more or less influenced by the adjoining precipitation gauges, the distance between the point and adjoining gauges has been used to adjust the magnitude of the precipitation as a modifier. The closer the gauge, the stronger the influences to a point can be expected. The mean annual precipitation pattern derived from miscellaneous sources (Scarf, 1972) has been used to adjust the precipitation at the point as another modifier:

$$R(y) = \sum_j \left\{ \frac{g[D(y, x_j)]}{\sum_i g[D(y, x_j)]} \right\} R(x_j) \frac{A(y)}{A(x_j)} \tag{1}$$

where $A(\ )$ is the mean annual precipitation at a point, $R(\ )$ is the daily precipitation at a point, y is the prediction point, and x_j is a precipitation gauge; $D(y, x)$ is the distance between the predictive point and the precipitation gauges. For this study, the inverse distance was used as the weighting function. The gauges with missing data were excluded from the weighting calculation. Therefore, the predicted precipitation at a point will be influenced by the 18 gauges around the catchment and the distribution of annual precipitation.

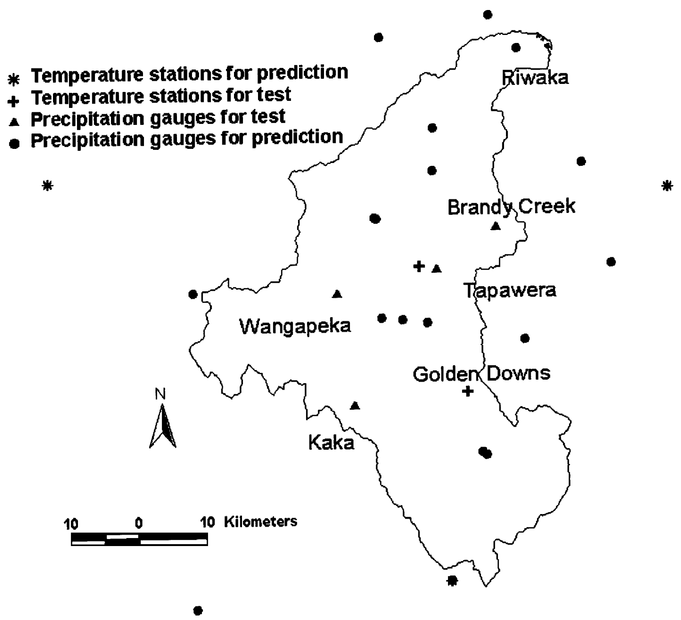

Figure 2. The precipitation and temperature gauges around the Motueka catchment

A 25 km spatial filter has been used to eliminate the influence of distant gauges. The distance of 25 km was chosen after trial-and-error analyses. In summary, the method favours the gauges that are near the given point of prediction (within 25 km), and more-distant gauges were excluded from the weighting calculation.

As part of the validation exercise, four other independent precipitation gauges with short historical records were selected for testing the daily precipitation predictive model (Figure 2).

Daily temperature prediction

The catchment elevation varies from sea level to 1850 m; therefore, a temperature lapse rate was used to predict the daily maximum and minimum temperature in each sub-basin. The calculation is

$$T = (1/n)\sum T_i \left(1 - \frac{H_i}{1000}\right) \mathrm{lr} \left(1 - \frac{H}{1000}\right) \mathrm{lr} \tag{2}$$

where T (°C) is the daily maximum and minimum temperature at a point, n is the number of temperature stations used for prediction, T_i (°C) is the daily maximum and minimum temperature at station i, H (m) is the elevation at the point of prediction, H_i (m) is the elevation at the temperature station i, and lr (°C km^{-1}) is the temperature lapse rate.

In this study, three temperature stations with long historical records were selected for daily temperature prediction. Temperature lapse rates of 7 °C km^{-1} and 4 °C km^{-1} respectively were used for the summer (November–March) and winter (April–October) periods (Barringer, 1989).

As part of validation, two other independent temperature gauges were selected for testing the daily temperature predictive model (Figure 2).

Runoff

Surface runoff is estimated by the Soil Conservation Service curve number method (Soil Conservation Service, 1972). The curve number varies non-linearly from condition I (dry) at wilting point to condition III (wet) at field capacity, and approaches 100 at saturation.

Three antecedent moisture conditions were defined by Soil Conservation Service (1972) and the curve number changes in response to the change of soil antecedent moisture conditions in the SWAT model: I, dry (wilting point); II, average moisture; III, wet (field capacity). Two-thirds of the Motueka catchment is steep country, with slopes exceeding 27%, and the typical curve numbers in antecedent moisture condition II are assumed to be appropriate for 5% slopes. Therefore, an algorithm (Williams, 1995) was used to adjust the curve number for different slopes. Usually, steeper slopes result in an increase in curve number.

$$CN_{2s} = \frac{CN_3 - CN_2}{3}[1 - 2\exp(-13{\cdot}86\text{slp})] + CN_2 \tag{3}$$

where, CN_{2s} is the curve number in moisture condition II adjusted for slope, CN_3 is the curve number moisture condition III for the default 5% slope, CN_2 is the moisture condition II curve number for the default 5% slope, and slp is the average percentage slope of the sub-basin. CN_3 is calculated as:

$$CN_3 = CN_2 \exp[0{\cdot}00\,673(100 - CN_2)] \tag{4}$$

Soil and root-zone infiltration and baseflow

A storage routing technique is used to predict infiltration through each soil layer (up to 10 layers) in the root zone. Downward flow occurs when field capacity of a soil layer is exceeded if the layer below is not saturated. The downward flow is governed by the saturated conductivity of the soil layer. A kinematic storage routing technique that is based on saturated conductivity is used to calculate lateral subsurface flow simultaneously with percolation.

A shallow aquifer storage recharged by the percolation from the bottom of the root zone is incorporated. Baseflow is allowed to enter the channel reach only if the amount of water stored in the shallow aquifer exceeds a threshold value defined through a calibration process.

Evapotranspiration

Evapotranspiration is the primary mechanism by which water is removed from a catchment. Three options for estimating PET are included in the model: Penman–Monteith method (Monteith, 1965); Priestley–Taylor method (Priestley and Taylor, 1972); and the Hargreaves method (Hargreaves and Samani, 1985). The Penman–Monteith method requires solar radiation, air temperature, relative humidity and wind speed. The Priestley–Taylor method requires solar radiation, air temperature and relative humidity. The Hargreaves method requires daily air temperature as input. The Hargreaves method (Hargreaves and Samani, 1985) was selected to calculate the PET throughout the catchment.

The model computes evaporation from soils and plants separately. Actual evaporation is first computed from any wet leaf evaporation (canopy interception). The canopy storage capacity (CSC) is the maximum water intercepted by vegetative surfaces, where it is held and made available for evaporation. The water that is held in canopy storage varies from day to day as a function of the leaf area index. Soil water evaporation is estimated as a function of PET and leaf area index.

CSC varies largely depending on the age and stand density of the vegetation types. The averaged CSCs for forest species in New Zealand range from 1 to 3 mm, with scrub at about 2 mm and tussock grassland at 0·6 mm (Rowe, 1983; Fahey *et al.*, 2001). Table II lists the CSCs used for land cover types in the Motueka catchment.

RESULTS AND DISCUSSION

The SWAT calibration and validation procedure followed several steps (Figure 3). First, the predicted daily precipitation and temperature were tested using the independent gauges in the catchment. The PET calculation in SWAT was also validated using field measurement and published data at different sites and temporal scales.

Table II. Canopy storage capacities for land cover types in the Motueka catchment

Land cover	Canopy storage capacity (mm)
Exotic forest	2·8
Indigenous forest	2·3
Scrub	1·5
Tussock	0·6
Pasture	1·8
Bare land and water	0

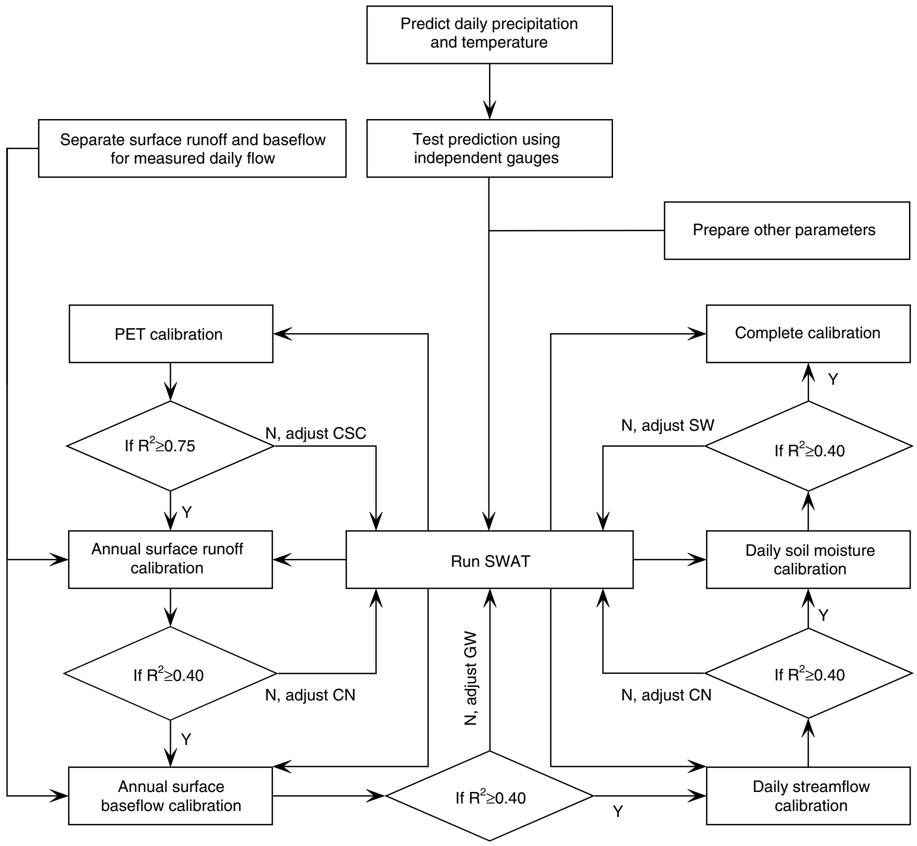

Figure 3. Flow calibration procedure for SWAT in the Motueka catchment

Second, the computer hydrograph separation program HYSEP (Sloto and Crouse, 1996) was used to derive baseflow and surface runoff from measured flow data in each subcatchment. The HYSEP-derived mean annual surface runoff and baseflow were used to calibrate the simulated surface runoff and baseflow respectively. The period 1990–94 was used for daily streamflow calibration. The daily streamflow from the whole catchment at Woodstock was calibrated to reflect contributions from annual surface runoff and baseflow. After calibration of the daily streamflow at Woodstock, the daily streamflow from subcatchments was fine-tuned (Figure 3). The period 1995–2000 was used for validation. Third, soil moisture measured in the field was used for model calibration (Figure 3). The multi-site calibration started with the subcatchments farthest upstream, followed by the next downstream subcatchment, with the total catchment above Woodstock being the final area used for calibration (Figure 4). Soil water content was compared only in Waiwhero subcatchment.

In presenting the results, the Nash–Sutcliffe coefficient (Nash and Sutcliffe, 1970) and R^2 values have been used. A Nash–Sutcliffe coefficient value of unity represents a perfect match, and smaller values represent poorer results.

Daily precipitation and temperature

In view of the high heterogeneity and variability of precipitation in the catchment, the daily precipitation predictive model generally had an acceptable performance. Owing to the regular spacing and good network of reference gauges (Figure 2), the Nash–Sutcliffe coefficient at three precipitation test sites exceeded 0·65 (Table III) and more than 62% of the precipitation variance was explained. However, at Kaka, the model performed poorly, with a Nash–Sutcliffe value of 0·36 (Figure 2). This was ascribed to the dissected terrain at Kaka, and its relatively distant geographical position from other precipitation gauges. Precipitation at the meso-scale in the South Island of New Zealand generally results from topographic modification of synoptic-scale airflow dominated by westerly winds (Mosley and Pearson, 1997). Rugged terrain results in highly variable precipitation and difficulties in spatially estimating the precipitation. Geographically, the precipitation predictive model performed well in the lower part and poorly in the upper part of the Motueka catchment due to the difference in terrain and the spacing and network of the precipitation gauges.

The predictive model assumes the precipitation on a particular day within the catchment always coincides with the annual precipitation pattern; thus, the daily prediction was significantly modified by the annual precipitation pattern. Although a good accuracy in annual temporal scale was expected, the prediction uncertainty in the daily scale increased.

Temperature varies with altitude, and high variability is expected in the Motueka catchment owing to the complex terrain. The predicted daily maximum and minimum temperatures were in agreement with those measured at two test gauge sites, with the Nash–Sutcliffe coefficients exceeding 0·83, and more than 90% of the variance in temperature was explained (Table IV). This shows that, although only three gauges with irregular spacing were used, the model performed well in predicting daily maximum and minimum temperatures. The use of an interpolation based on known lapse rates does, therefore, account for temperature variability within the catchment.

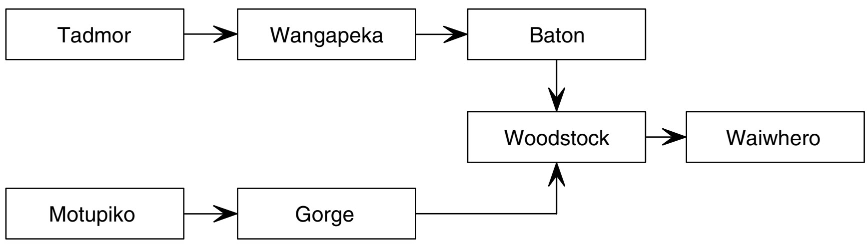

Figure 4. Multi-site calibration procedure in Motueka catchment

Table III. Predictive model test results for daily precipitation

No.	Test gauge	Altitude (m)	Duration of test	Nash–Sutcliffe	R^2
1	Tapawera	160	2 Aug 1992–31 Dec 2001	0·65	0·62
2	Wangapeka	240	1 Jan 1989–1 Nov 1996	0·70	0·71
3	Brandy Creek	222	1 Jan 1989–24 Dec 2001	0·65	0·65
4	Kaka	402	1 Jan 1989–31 Mar 1998	0·36	0·45

Table IV. Daily maximum and minimum temperature prediction test

Test gauge	Altitude (m)	Duration of test	Nash–Sutcliffe		R^2	
			Max.	Min.	Max.	Min.
Tapawera	160	1 Jan 1977–31 Jul 1987	0·83	0·84	0·92	0·91
Golden Downs	274	1 Jan 1974–31 Mar 1980	0·86	0·89	0·93	0·89

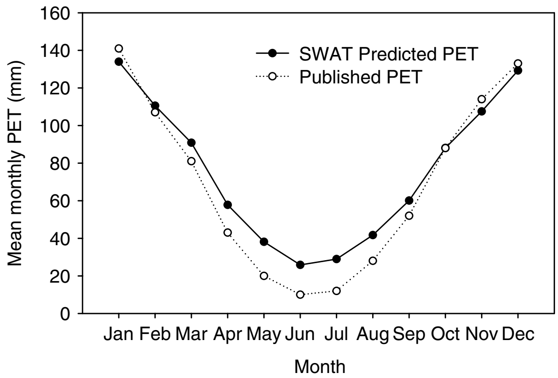

Figure 5. Hargreaves-predicted PET compared with the published data (New Zealand Meteorological Service, 1986) at Riwaka ($R^2 = 0{\cdot}99$)

PET

The Hargreaves method was selected to calculate PET throughout the catchment because some data were unavailable and the method's PET estimates were in agreement with the data from several sources and time scales. At Riwaka (Figure 2), the Hargreaves method predicted an annual mean PET of 844 mm in the 11-year period 1990–2000, consistent with the 829 mm published by the New Zealand Meteorological Service (1986) for 1965–83. In contrast, the other two PET methods, i.e. Penman–Monteith and Priestley–Taylor, produced a much lower PET (540 mm) and, therefore, were not used for model calibration. The poor performance of these is due to the interpolation of the input variables from only one site, actually outside the catchment. The predicted monthly PET by the Hargreaves method was also in agreement with the published PET data (Figure 5), with an R^2 of 0·99 at Riwaka.

A daily PET estimate for 1996–2000 based on monitored hourly meteorological data at central Moutere (near Woodstock) using a Penman–Monteith method recommended by the Food and Agriculture Organization (Allen *et al*., 1998) was also used for comparison. The Hargreaves method predicted an annual mean PET of 803 mm at this site, which was 34 mm higher than the 769 mm from the hourly 'measured' data calculation. The mean monthly PET showed a high correlation (Figure 6) between the Hargreaves predicted PET and the estimate based on the hourly meteorological data. Similarly, the predicted daily PET pattern also showed good agreement with the calculation based on the hourly data, with an R^2 of 0·78 (Figure 7).

The results clearly show that the Hargreaves method, based on the currently available meteorological data, had the best PET estimation at the yearly, monthly, and daily time scales for PET prediction in the Motueka catchment.

Baseflow and surface runoff

The R^2 for predicted annual baseflow against the hydrograph-separated baseflow from measured flow data in 11 years at Woodstock was 0·79 (Figure 8). This showed the model had a good performance in baseflow modelling over the whole catchment. However, the R^2 varied from 0·46 to 0·90 in the subcatchments (Table V).

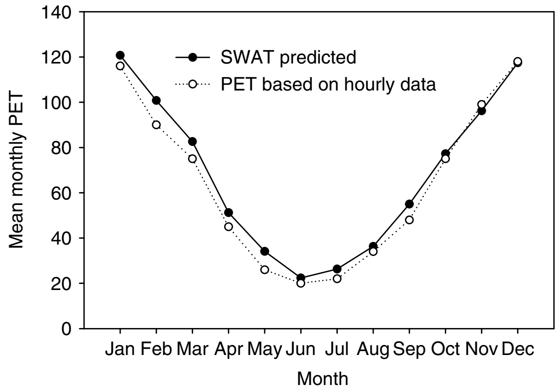

Figure 6. Hargreaves-predicted PET against calculated PET based on the hourly data at central Moutere ($R^2 = 0{\cdot}99$)

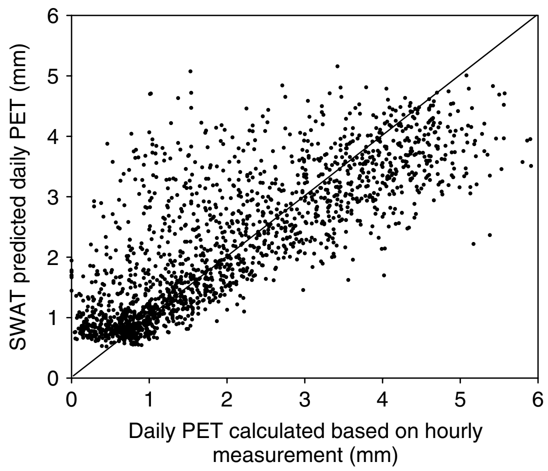

Figure 7. Hargreaves-predicted PET against calculated daily PET from 1996 to 2000 at central Moutere ($R^2 = 0{\cdot}78$)

A similar pattern (Figure 8), which showed a predicted annual baseflow somewhat higher than hydrograph-separated baseflow, was observed in all subcatchments except upper Motueka catchment at Gorge.

Annual total water yield

All the subcatchments are nested in the Motueka catchment upstream of Woodstock; therefore, the hydrological response at Woodstock depends on the combined behaviour of upstream subcatchments. The predicted annual total water yield matched the measured value well at Woodstock, with an R^2 of 0·91 (Figure 9g), and generally had an acceptable accuracy in five of the six subcatchments (Baton, Stanley Brook, Tadmor, Wangapeka and Motupiko) with R^2 ranging from 0·64 to 0·95 (Figure 9a–g). However, in the upper Motueka (Gorge gauging site), a significant discrepancy between the predicted and measured water yields was observed (Figure 9a).

Predicted and measured annual water yields were reasonably, but not highly, consistent in the Baton and Wangapeka subcatchments, reflecting precipitation variability in the high-altitude and rugged terrain. At Motueka Gorge, the gap between predicted annual water yield and measured water yield clearly indicates a poor predicted precipitation, and this deficiency in precipitation prediction consequentially contributed to the lower predicted annual water yield in the whole Motueka catchment below Woodstock (Figure 9g). This has

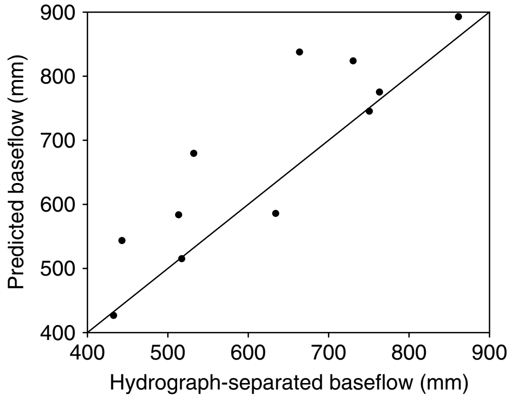

Figure 8. Predicted against the hydrograph-separated annual mean baseflow over 11 years at Woodstock

Table V. Predicted and hydrograph-separated baseflow for subcatchments

No.	Subcatchment	R^2
1	Upper Motueka at Gorge	0·70
2	Motupiko	0·75
3	Tadmor	0·90
4	Baton	0·46
5	Wangapeka	0·58
6	Stanley Brook	—[a]
	Motueka at Woodstock	0·79

[a] Not enough data to calculate R^2 at Stanley Brook.

been recognized as a deficiency in precipitation prediction, and another rain gauge has been installed in the upper Motueka area.

Daily streamflow

The Nash–Sutcliffe coefficient at Woodstock for daily streamflow was 0·78 during the calibration period, and the value varied from 0·36 to 0·61 in the subcatchments (Table VI). For the validation period, the coefficient was 0·72 at Woodstock, with a range from 0·35 to 0·57 in the subcatchments. R^2 was 0·82 for the calibration period and 0·75 for the validation period for daily streamflow at Woodstock. R^2 exceeded 0·5 for the subcatchments, except the upper Motueka at Gorge site during the validation period (Table VI). This was expected, owing to the insufficient predicted precipitation shown in the annual water yield comparison.

The daily streamflow was predicted well in the Motueka catchment at Woodstock (Figure 10). For subcatchments Wangapeka, Tadmor and Stanley Brook, the model generally had an acceptable efficiency, but the model performed poorly in other parts of the catchment, such as Baton and upper Motueka at Gorge.

As the predicted annual water yield indicated, the daily streamflow prediction verified that precipitation in Baton, Wangapeka, and upper Motueka above Gorge was highly variable. The difficulty in predicting the spatial variability of precipitation appeared to be the main contributor to the model performance in daily streamflow prediction. Obviously, prediction biases from upstream subcatchments were self-compensating at the larger catchment scale, and thus the model performed better at Woodstock.

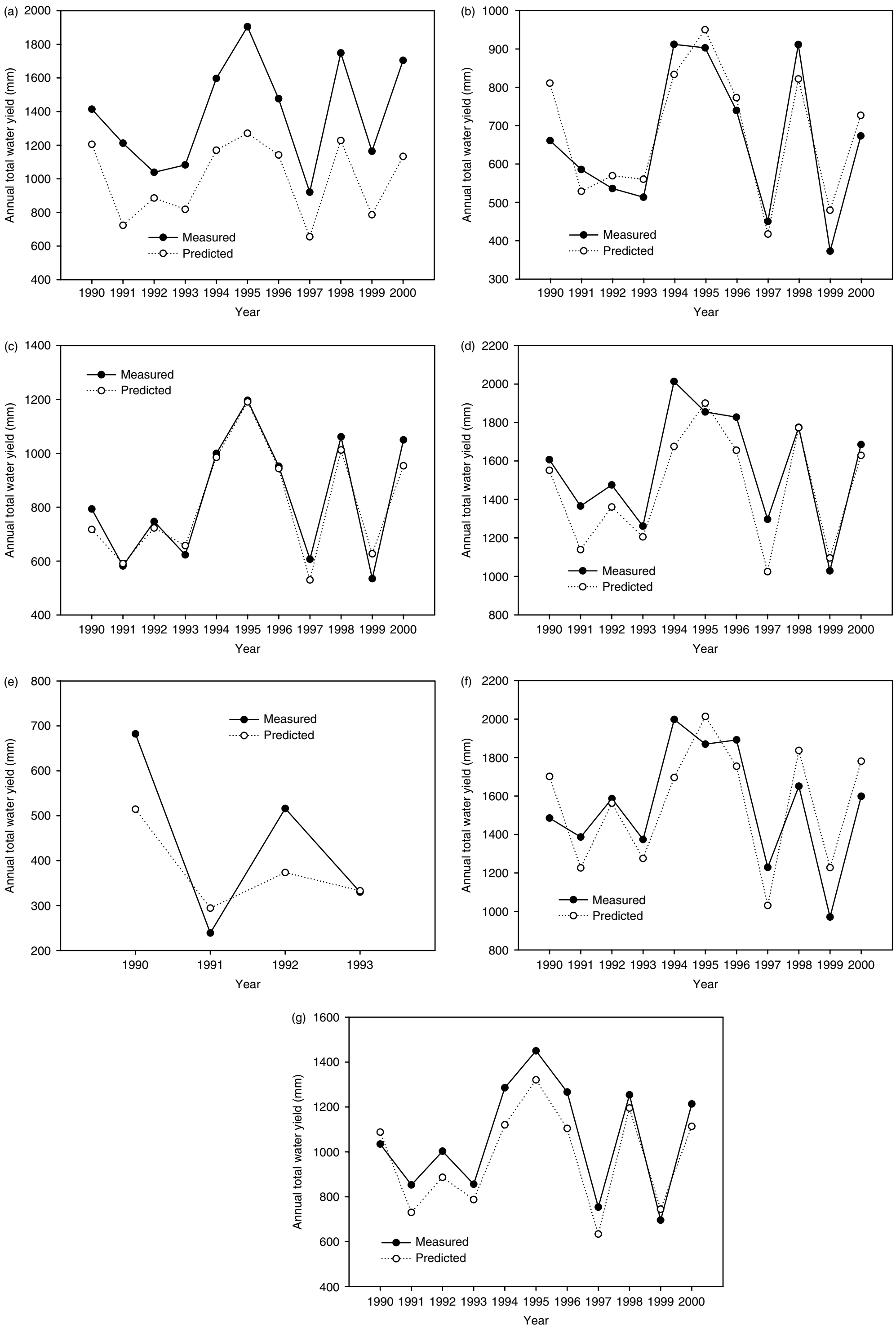

Figure 9. Predicted annual water yield and that measured at Woodstock and in different subcatchments: (a) Gorge; (b) Motupiko; (c) Tadmor; (d) Wangapeka; (e) Stanley Brook; (f) Baton; (g) Woodstock

Table VI. Daily streamflow calibration and validation results

Subcatchment	Period	Nash–Sutcliffe coefficient	R^2
Upper Motueka at Gorge	Calibration	0·42	0·52
	Validation	0·41	0·41
Motupiko	Calibration	0·40	0·54
	Validation	0·57	0·61
Tadmor	Calibration	0·61	0·61
	Validation	0·55	0·56
Wangapeka	Calibration	0·60	0·62
	Validation	0·51	0·53
Baton	Calibration	0·36	0·60
	Validation	0·35	0·51
Stanley Brook	Calibration	0·59	0·65
	Validation	No records available for validation	
Motueka at Woodstock	Calibration	0·78	0·82
	Validation	0·72	0·75

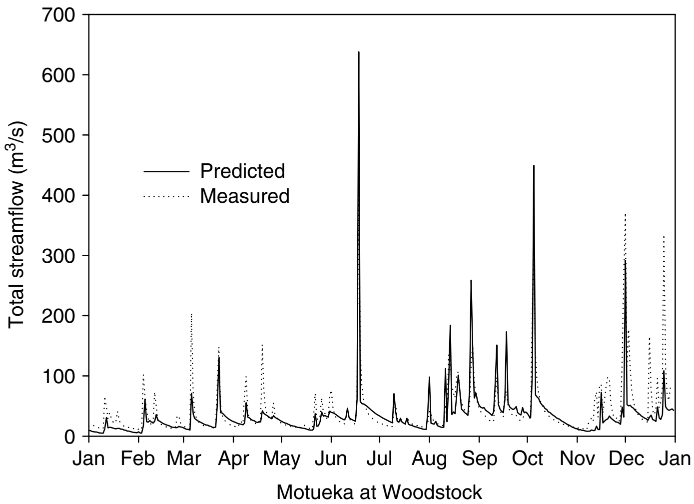

Figure 10. Predicted daily streamflow against the measured at Woodstock in 1997

Soil moisture

Soil moisture was only available for the small Waiwhero subcatchment, area 3·9 km^2. A neutron probe was used to measure soil moisture in the field on a weekly basis during the late 1990s. From the catchment ridge to hill-slope bottom, two soil transects were marked on one flank of the slope, and one transect on another shorter side. Six soil-moisture access tubes along each soil transect and one from a sampling site near the outlet were recorded on a fortnightly basis. The soil moisture from the 19 sites has been averaged to represent the areal soil moisture for the whole Waiwhero subcatchment. This intensive sampling strategy, to some extent, resolved the problem raised by Beven (1989), that average areal soil moisture predicted by a model cannot be used for comparison with a 'point' neutron probe measurement.

The soil moisture samples were recorded in a period from 30 May 1997 to 22 July 1999, and the precipitation was measured simultaneously. The measured soil moisture varied from very dry (0 mm) to more than field capacity (162·1 mm, Figure 11). However, SWAT predicted that soil moisture would remain wet, ranging from 106 mm to 142 mm, in the same period. This significant discrepancy can be explained partly by the

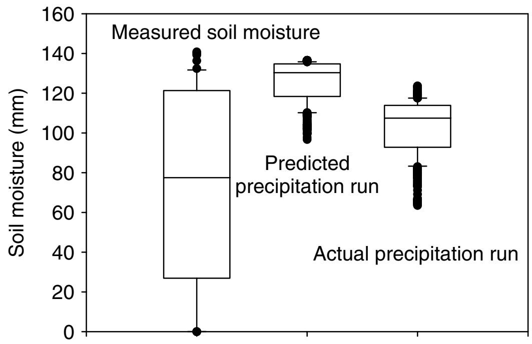

Figure 11. Soil moisture comparison between the measured and the predicted (box-and-whisker figure shows the range of the moisture distribution; the line within the box shows the mean moisture value)

Table VII. Predicted precipitation and that measured in the Waiwhero catchment

	Total days (days)	Total precipitation (mm)	Wet days (days)	Mean precipitation per event (mm)	Precipitation standard deviation (mm)
Predicted	783	2688	486	5·5	7·7
Measured	783	2327	241	9·7	9·0

fact that the predicted wet days were twofold more than the measured (Table VII), and about 360 mm more precipitation was predicted in the monitoring period. The actual precipitation, which tended to be of high intensity and short duration, was different to predicted precipitation. A new SWAT run with actual measured precipitation at the site suggested the error derived from the difference between predicted and measured precipitation accounted for around an average 23 mm (about 50%) of the extra soil water storage that the model had predicted previously in the period. The remaining average 28 mm (another 50%) discrepancy was likely to be a result of poor soil property representation in the catchment.

Discussion

A variety of factors may be responsible for the model errors when comparing measured streamflow data with simulated output. These include model parameterization uncertainty (spatial variability in climate, soil and land use), flow measuring uncertainty, errors due to sampling strategies, and errors or oversimplifications inherent in the model structure (Refsgaard and Storm, 1996). It has been shown that the accuracy of the daily precipitation prediction model depends significantly upon the precipitation gauge numbers and their location. The Wangapeka, upper Motueka at Gorge, and Baton subcatchments are in areas of highly variable precipitation. However, no precipitation gauge records were available to record from the upper parts of the Baton and upper Motueka subcatchments, which resulted in poor streamflow prediction for Baton and upper Motueka at Gorge

It is likely that even if different individual parameters were optimal in the situations where they were determined, bringing their values together from different sources is no guarantee that they will give good results as a set in a new set of circumstances (Beven, 2001). This reflects the problem of equifinality in PDHMs. When SWAT was applied to the larger Motueka catchment it predicted annual and daily streamflows with an adequate degree of accuracy. However, analysis of the prediction of separate internal hydrological processes and subcatchments showed somewhat poorer predictive ability, suggesting that the result was compensating between differing factors at the larger scale. Theoretically, compensations in the model results derive hydrologically from the water balances in the various pathways and storage (evaporation, streamflow, surface runoff, soil moisture and groundwater), or geographically from the contributions of different areas (hill slopes, variable source areas, sub-basins, and subcatchments) to streamflow.

As emphasized by Beven (2001), limited measurements and poor understanding of subsurface processes in particular will result in equifinality. In our modelling work, although a significant contributor to model errors was verified (spatial variability of the precipitation), the calibrated parameters may be just one set of parameterizations that can produce acceptable results. However, owing to the complexity in hydrological processes, a complex mixture of land use and soil properties, and a nested distribution of subcatchments, our work successfully yielded an intersection of alternative parameterizations after the multi-site and multi-variable calibration approach was used. This greatly reduced uncertainties from equifinality problems during parameterization. The use of an integrated multi-variable and multi-site calibration and validation method improved the model calibration and validation and highlighted the areas and the hydrological processes requiring greater calibration effort (e.g. understanding precipitation distribution and soil moisture change).

The other issue arising in this study was the multiple temporal-scale tests. As the PET and streamflow prediction showed, the daily prediction in PET and streamflow did not match the measured data as well as the annual prediction did. The monthly PET prediction by SWAT at central Moutere had a high agreement with that estimated using hourly measurements, but a 34 mm difference still occurred between annual predicted and measured PET. Thus, a fine-scale or multi-temporal-scale validation is strongly recommended.

CONCLUSIONS

A proposed calibration approach integrating multiple internal variables and multiple sites was used to develop a SWAT model application at a large scale for the Motueka catchment. Daily precipitation and temperature were predicted using interpolation-based, separated models and available meteorological data. The hydrological components of the SWAT model, such as PET, water yield, streamflow and baseflow, were calibrated and validated at whole-catchment scale and for six subcatchments of the Motueka catchment. This multi-variable and multi-site calibration and validation approach resulted in more realistic parameter values across both the hydrological processes and the geographic areas, and highlighted the areas (e.g. upper Motueka) and the hydrological processes (e.g. soil moisture) requiring greater calibration effort. However, the spatial variability of precipitation could be better represented, this contributing significantly to model errors.

Given the high spatial variability of the precipitation, the integrated calibration and validation process showed that SWAT had an acceptable hydrological performance in the Motueka catchment. This integrated multi-variable and multi-site calibration and validation method also produced more realistic input parameters for the SWAT model, reducing the errors from the potential equifinality or the non-uniqueness problem in PDHMs. Although the conclusions drawn from such work are site and model component dependent, the research, nevertheless, has an important role to play in spatially calibrating and validating a hydrological model in a large catchment.

This work should be considered as a first step to developing a model useful for catchment planning, and much work still needs to be done on model verification. In the next phase of this research, nutrient and sediment variables will be included for spatially calibrating and validating the SWAT model against the regularly monitored data from different tributaries. A long-term research plan in the Motueka catchment will continuously promote knowledge of the model calibration and validation, and also improve the prediction of our testing scenarios.

ACKNOWLEDGEMENTS

We gratefully acknowledge the funding for this study from Landcare Research, New Zealand, and the Foundation for Research, Science and Technology (C09X0305). We would also like to thank two reviewers for their valuable comments on the manuscript.

REFERENCES

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

Allen RG, Pereira LS, Raes D, Smith M. 1998. *Crop evapotranspiration—guidelines for computing crop water requirements*. FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization of the United Nations, Rome.

Anderton S, Latron J, Gallart F. 2002. Sensitivity analysis and multi-response, multi-criteria evaluation of a physical based distributed model. *Hydrological Processes* **16**: 333–353.

Arnold JG, Allen PM. 1996. Estimating hydrologic budgets for three Illinois watersheds. *Journal of Hydrology* **176**: 55–77.

Arnold JG, Srinivasan P, Muttiah RS, Williams JR. 1998. Large area hydrologic modeling and assessment. Part I. Model development. *Journal of the American Water Resources Association* **34**: 73–89.

Barringer J. 1989. A variable lapse rate snowline model for the Remarkables, Central Otago, New Zealand. *Journal of Hydrology (New Zealand)* **28**(1): 33–46.

Bergstrom S, Lindstrom G, Pettersson A. 2002. Multi-variable parameter estimation to increase confidence in hydrological modeling. *Hydrological Processes* **16**: 413–421.

Beven KJ. 1989. Changing ideas in hydrology: the case of physically based models. *Journal of Hydrology* **105**: 157–172.

Beven KJ. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* **16**: 41–51.

Beven KJ. 1996. A discussion of distributed modelling. In *Distributed Hydrological Modelling*, Refsgaard JC, Abbott MB (eds). Kluwer: Dordrecht; 255–278.

Beven KJ. 2001. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Science* **5**(1): 1–12.

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

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

Beven KJ, Hornberger GM. 1982. Assessing the effect of spatial pattern of precipitation in modelling streamflow hydrographs. *Water Resource Bulletin* **18**: 823–829.

Chittenden ET, Hodgson L, Dodson KJ. 1966. *Soils and agriculture of Waimea Country, New Zealand.* Soil Bureau Bulletin 30, DSIR, Wellington.

Duan Q, Sorooshian S, Gupta VK. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28**: 1015–1031.

Fahey B, Watson A, Payne J. 2001. Water loss from plantations of Douglas-fir and radiata pine on the Canterbury Plains, South Island, New Zealand. *Journal of Hydrology (New Zealand)* **40**: 77–96.

Fenemor AD. 1988. *A three-dimensional model for management of the Waimea Plains aquifers, Nelson*. Publication 18 of the Hydrology Centre, Christchurch.

Fohrer N, Haverkamp S, Eckhardt K, Frede H-G. 2001. Hydrologic response to land use changes on the catchment scale. *Physics and Chemistry of the Earth, Part B: Hydrology Oceans and Atmosphere* **26**: 577–582.

Grayson RB, Moore ID, McMahon TA. 1992a. Physically based hydrologic modelling I. A terrain-based model for investigative purposes. *Water Resources Research* **28**: 2639–2658.

Grayson RB, Moore ID, McMahon TA. 1992b. Physically based hydrologic modelling II. Is the concept realistic? *Water Resources Research* **26**: 2659–2666.

Gupta VK, Sorooshian S, Yapo PO. 1998. Toward improved calibration of hydrologic model: multiple and noncommensurable measures of information. *Water Resources Research* **34**: 751–763.

Gupta VK, Bastidas LA, Sorooshian S, Shuttleworth WJ, Yang ZL. 1999. Parameter estimation of a land surface scheme using multi-criteria methods. *Journal of Geophysical Research* **104**: 19 491–19 503.

Hamlin MJ. 1983. The significance of rainfall in the study of hydrological processes at basin scale. *Journal of Hydrology* **65**: 73–94.

Hargreaves GH, Samani ZA. 1985. Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture* **1**: 96–99.

Monteith JL. 1965. Evaporation and the environment. In *The State and Movement of Water in Living Organisms, XIXth Symposium.* Cambridge University Press: Swansea; 205–234.

Mosley P, Pearson C. 1997. Introduction: hydrological extremes and climate in New Zealand. In *Floods and Droughts: The New Zealand Experience*, Mosley P, Pearson C (eds). New Zealand Hydrological Society: Wellington North; 1–14.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, part I—a discussion of principles. *Journal of Hydrology* **10**: 282–290.

New Zealand Meteorological Service. 1986. *Summaries of water balance data for New Zealand Stations*. New Zealand Meteorological Service Miscellaneous Publication 189. Wellington.

Priestley CHB, Taylor RJ. 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review* **100**: 81–92.

Refsgaard JC. 1997. Parameterization, calibration and validation of distributed hydrological models. *Journal of Hydrology* **198**: 69–97.

Refsgaard JC, Storm B. 1996. Construction, calibration and validation of hydrological models. In *Distributed Hydrological Modelling*, Abbott MB, Refsgaard JC (eds). Kluwer Academic Press: The Netherlands; 23–26.

Rowe LK. 1983. Rainfall interception by an evergreen beech forest, Nelson, New Zealand. *Journal of Hydrology* **66**: 143–158.

Scarf F. 1972. Mapping annual surface water resources of the hydrological regions of Nelson, New Zealand. *Journal of Hydrology (New Zealand)* **11**: 105–126.

Shah SMS, O'Connell PE, Hosking JRM. 1996. Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. *Journal of Hydrology* **175**: 89–111.

Sloto RA, Crouse MY. 1996. *HYSEP: a computer program for streamflow hydrograph separation and analysis*. Water-Resources Investigations Report 96–4040, US Geological Survey, Reston, VA.
Soil Conservation Service. 1972. Hydrology. In *National Engineering Handbook*, Section 4. USDA Soil Conservation Service: Washington, DC.
Sorooshian S, Gupta VK. 1995. Model calibration. In *Computer Models of Watershed Hydrology*, Singh VP (ed.). Water Resources Publications: Highlands Ranch, CO; 23–68.
Srinivasan R, Ramanarayanan TS, Arnold JG, Bednarz ST. 1998. Large area hydrologic modeling and assessment part II: model application. *Journal of American Water Resource Association* **34**: 91–101.
Wang QJ. 1991. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resources Research* **27**: 2467–2471.
Williams JR. 1995. The EPIC model. In *Computer Models of Watershed Hydrology*, Singh VP (ed.). Water Resources Publications: Highlands Ranch, CO; 909–1000.

