Research Article

Data-Based Mechanistic Rainfall-Runoff Modelling for a Large Monsoon-Dominated Catchment in Thailand

Sukanya Vongtanaboon[1*], Han She Lim[2,3] and Keith Richards[3]

[1]*Faculty of Science and Technology, Rajabhat Phuket University, Phuket, Thailand*
[2]*Department of Geography, National University of Singapore, Singapore*
[3]*Department of Geography, University of Cambridge, Cambridge*
[*]*Corresponding author. E-mail address: vongtanaboon@yahoo.com*

Received July 7, 2007; Acceptd March 25, 2008

Abstract

Hydrological modelling for water resource and flood management in large monsoon-dominated sub-tropical catchments has not been the subject of extensive research and it is not clear what the appropriate model structures and data requirements may be. High degrees of seasonality, limited data availability, rapidly changing hydrological regimes as a result of land use change and climate variability and a lack of complete understanding of the details of the physical hydrology in these regimes and regions all contribute to this situation. This paper uses Data Based Mechanistic (DBM) modelling methods to explore the hydrology of the 3,853 km^2 Mae Chaem catchment in northern Thailand, where there is an unusually rich database of runoff and rainfall data. This is used to examine the appropriate model structure and parameter values in DBM models and the effects of using the available rainfall and runoff data in a range of different ways. Rainfall data are area-weighted using Theissen polygons, within which altitude adjustment is effected on the basis of evidence for an increase of about 0.5 mm of rain per rainday for each 100 m increase in elevation above 1000 m, in the monsoon season only. The model structure suggests a second-order model and the parameter values seem to be rather stable when higher quality rainfall data are used. Furthermore, it is possible to maintain reliable flow simulations by cascading a series of runoff prediction regression models that predict a downstream flow from an upstream flow and the incremental rainfall between gauging stations.

***Key Words*:** Data-based mechanistic modelling; Monsoon; Rainfall-runoff modelling; Semi-distributed rainfall and runoff data; Thailand

Introduction

Flood disasters occur relatively frequently in the monsoon climate of Thailand, resulting in loss of life and significant economic damage. Furthermore, given that the consequences of flooding are serious and recurrent, the government has initiated a plan to prevent and/or mitigate flood disasters, based on modelling rainfall-runoff processes during the monsoon flood season, and establishing flood forecasting and warning systems. Currently, flood warning is based largely on empirical (but effective) rules of thumb identifying the upstream water levels associated 6-12 hours later with critical flood levels in downstream urban areas. Further development of the methodology for rainfall-runoff modelling in Thai river basins is likely to continue to use these methods, but will be supplemented by spatially-distributed rainfall and runoff data inputs, and may involve integration of rainfall radar data, and will ultimately involve forecasting both the generation and routing of runoff, and the subsequent occurrence of downstream flooding.

Objectives

As a contribution to this project, this paper explores the use of data-based mechanistic (DBM) modelling for the large monsoon-dominated Mae Chaem catchment, in Chiang Mai province, northern Thailand. The aims are to identify an optimal and efficient model structure; to assess the stability of parameter values from one year's monsoon to the next; to assess the stability of parameter values in relation to variation in the quality of input rainfall data; to examine the physical inferences that can be made from the parameter values; to investigate the potential for a semi-distributed use of DBM modelling; and to explore the potential of the DBM modelling method in forecasting floods in the catchment.

The Mae Chaem River

The Mae Chaem is a major north-south oriented tributary of the Upper Ping River, which is the largest tributary of the Chao Phraya river system. It is composed of 3,853 km^2 of steep, originally forested terrain, is located about 70 km west of Chiang Mai, the northern provincial capital of Thailand, and is approximately bounded by latitudes 18^o 12' N to 19^o 8' N and longitudes 98^o 8' E to 98^o 30' E. The altitude of the catchment falls sharply from the highest point on its eastern watershed (Doi Inthanon, at 2565 m) to 273 m at the town of Amphoe Hot, the junction of the Mae Chaem with the Ping River. At about 10 km upstream from Amphoe Hot, the Mae Chaem is gauged at the Royal Irrigation Department (RID)-maintained P14 gauging station, which has a 50-year record. The 2001 flood in the Mae Chaem occurred at 0400 hrs on 13 August, with a peak stage of 7.98 m (283.78 msl), and a discharge of 794 m^{3}s^{-1} (extrapolated from the rating curve data updated annually by the Thai Royal Irrigation Department).

The Mae Chaem catchment is in the seasonal humid tropics, with a climatic regime dominated by the south-west monsoon. The rainfall has distinct wet and dry seasons, with the wet monsoon season typically beginning in April or May and lasting until October. The dry season, from November-December to March, has minimal rainfall, with monthly totals of zero being common. Average annual rainfall at Amphoe Mae Chaem, a town centrally located within the catchment, is 973.3 mm (based on 47 years of data). Figure 1 shows the seasonal rainfall regime for two rainfall stations in the catchment at different elevations; this indicates a strong orographic effect, with higher rainfall on the west and east watersheds. There is a strong dependency of rainfall on elevation above 1000m during the monsoon (Figure 2), probably reflecting the recurrent thermal and humidity characteristics of the SW monsoon air flow, and typical condensation levels in the air mass. There is also evidence of an increasingly strong ENSO signal in the year-to-year rainfall variability; 1997-1998 in particular coincided with an El Nino event

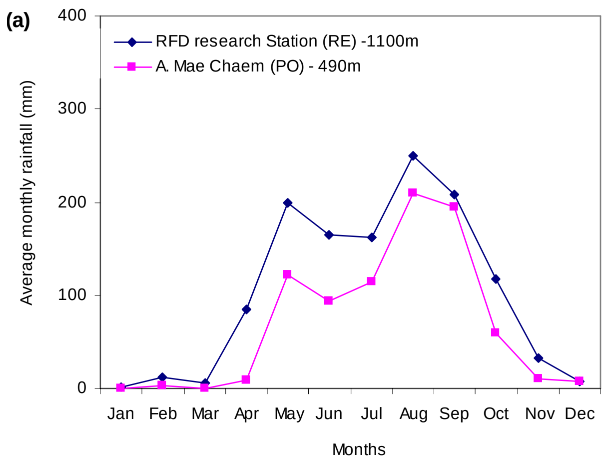


Figure 1 The seasonal rainfall regime for two rainfall stations at different elevations in the Mae Chaem catchment, north-western Thailand (averages of data for the period 1996-2002).

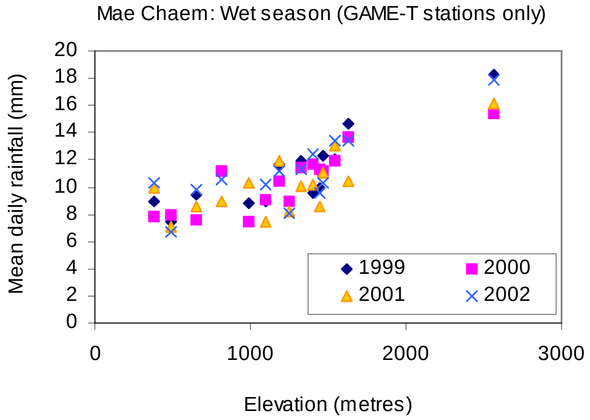


Figure 2 The strong dependency of rainfall on elevation above 1000m during the monsoon; data for the GAME-T raingauges, and for the years 1999-2002

with annual rainfall totals 65-75% of the levels in 1999-2002.

There is a range of rainfall data available in the catchment. Rainfall data are available from a rainfall station operated by the Royal Irrigation Department (RID) at Amphoe Mae Chaem (Figure 3). The Global Energy and Water Cycle Experiment (GEWEX) has established for its Asian Monsoon Experiment-Tropics (GAME-T) a raingauge network in the Mae Chaem watershed to determine rainfall in the mountainous the upper Chao Phraya river basin. Automatic tipping bucket rain gauges with pulse-count time-recording data loggers were installed at 15 sites (1 to 15 in Figure 3). Most of these gauges are at

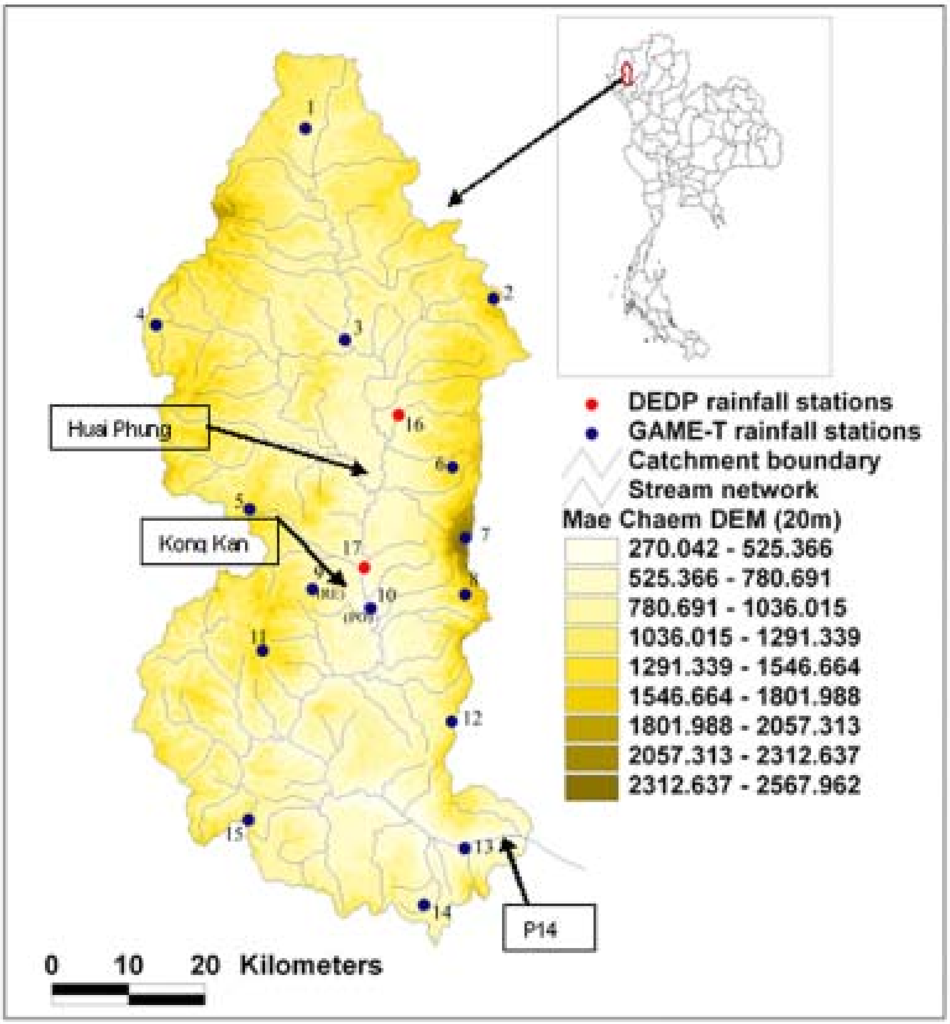


Figure 3 Rainfall stations for which data are available in the Mae Chaem catchment, Chiang Mai province (in addition, the three river gauging stations of Huai Phung, Kong Kan are also shown).

elevations above 1000m and daily data are available for 1998 to 2003 (Koichiro et al., 2001). Two more rainfall stations are operated by the Department of Energy Development and Promotion (DEDP), located near the Mae Mu and Kong Kan flow gauging stations and shown as 16 and 17 in Figure 3. The Royal Forestry Department also runs a rainfall station at its Research Station. These rainfall data for the Mae Chaem catchment have been corrected by Lim (2006) using Thiessen polygons to define area weights, further adjusted for the effect of altitude above 1000m as shown in Figure 2. Thus, area-weighted and elevation-corrected rainfall data from 2000 to 2002 (the period when all GAME-T gauges were operational) are available for the Mae Chaem catchment. In addition to the rainfall data, daily discharge data are available from 1955 to 2003 for the RID gauging station near the basin outlet, P14 (at 18.23N, 98.56E), and at upstream sites at Huai Phung (where the catchment area is 1203.6 km^2), and at Kong Kan (where it is 2157.8 km^2).

Previous DBM modelling for the Mae Chaem catchment (Vongtanaboon, 2004; Vongtanaboon and Chappell, 2004) has suggested that a first-order DBM transfer function model fits the Mae Chaem rainfall-runoff data reasonably well, as judged by the R_t^2 criterion (Table 1). However, the results showed that hydrograph recessions were not described well by this model. Furthermore, on examining the first-order model parameters, it appeared that these vary considerably from year to year, suggesting that this model is variable in its fit from one monsoon to the next. The first-order model time constant (TC) is estimated to be 7 days using a bilinear power model and 13 days using the storage model transformation. However, investigations by Lim (2006) suggest that rapid conversion of rainfall into runoff and efficient routing of floods occurs in the Mae Chaem catchment. The maximum discharge recorded in the last 25 years at the downstream P14 gauging station, in the flood of 13 August 2001, occurred only one day after the peak for this hydrograph was recorded at an upstream gauging station. This indicates that the catchment rainfall-runoff behaviour needs to be modelled using a higher-order transfer function capable of defining both fast flow and slow flow pathways.

Methodology

The Data-Based-Mechanistic (DBM) dynamic transfer-function model is a parsimonious modelling method that quantifies uncertainty associated with model characterizations and predictions, and permits interpretation of the resulting model structure in physical terms (Young et al., 1997; Young, 2001). Transfer function modelling can be used to relate discharge to rainfall, and the associated model

Table 1 A comparison of the fitted first- and second-order models, using both the bi-linear power and the storage transformation, and applied successively to annual data from single rainfall station for the period from 1998-2002.

Year	Bi-linear power transformation				Storage transformation			
	[1,1,0] model		[2,2,1] model		[1,1,0] model		[2,2,1] model	
	R_t^2	YIC	R_t^2	YIC	R_t^2	YIC	R_t^2	YIC
1998	0.7336	-8.3521	0.6939	-6.3331	0.7490	-8.4290	0.7461	-6.8101
1999	0.5119	-6.1294	0.6235	-4.0749	0.4189	-4.9820	0.5603	-3.3820
2000	0.5475	-6.8137	0.7555	-6.1822	0.5275	-6.4740	0.7025	-5.5204
2001	0.6920	-7.9060	0.7808	-7.0801	0.7421	-8.1736	0.7915	-6.8069
2002	0.8654	-8.6338	0.9129	-5.6299	overflow		0.8180	-3.3041

parameters may have physical attributes that can be interpreted in terms of hydrological processes. In a simple input-output system such as this, the general form of transfer function reduces (Young, 1998) to:

$$yt = f(u_t) + N_t + e_t = f(u_t) + \xi_t \quad (1)$$

where y_t is the observed time series, $f(u_t)$ captures the influence of input variables (u_t), N_t is a stochastic perturbation component (usually modelled as an AutoRegressive (AR) or AutoRegressive, Moving Average (ARMA) process); e_t is an irregular component, usually defined as a normally distributed Gaussian sequence with zero mean value and variance σ^2 (i.e. discrete-time white noise) and ξ_t represents a general stochastic input that accounts for the combined effects of other factors affecting y_t, such as noise, other stochastic inputs or model limitations. Here the transfer function is linear; non-linearity only resides in the definition of the input variables through the non-linear function $f(u_t)$.

A discrete-time transfer function can be written in the form:

$$y_t = \sum_{i=1}^{i=m} \frac{B_i(z^{-1})}{A_i(z^{-1})} f_i(u_{t-d}) + \xi_t \quad (2)$$

where z^{-1} is the backward shift operator, and $A_i(z^{-1})$, $B_i(z^{-1})$ are appropriately defined polynomials in these operators (Young, 1998). In the simplest single input case, the transfer function polynomials are defined as follows:

$$A_i(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_n z^{-n} \quad (3)$$

$$B_i(z^{-1}) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_m z^{-m} \quad (4)$$

and the order of the associated transfer function model is defined by the triad [n,m,d], where n is the order of the denominator in (2), m is that of the numerator, and d is an advective time delay (Young et al., 1997). Non-linearity in the rainfall-runoff response can be captured in a DBM model in one of two alternative sub-models: the ‘bilinear power model’ and the ‘storage model’. These are used to define the ‘effective rainfall’ (or the flow once water has passed through the non-linear parts of the catchment – such as the unsaturated zone), which is then normalised by the actual rainfall, to maintain the overall mass balance.

The bilinear power model uses the current discharge as an index of antecedent moisture status of the catchment, in the transformation:

$$u(t) = R(t)\, Q(t)^p \quad (5)$$

where u(t) is an effective rainfall, R(t) is rainfall, Q(t) is discharge and *p* is the power coefficient (Young and Beven, 1994). The storage model, on the other hand, uses a soil moisture index at each time step, and the effective rainfall is derived from:

$$u(t) = S(t)R(t) \quad (6)$$

where S(t) is a catchment storage index or catchment wetness/soil moisture index from:

$$S(t) = S(t-1) + \{R(t) - S(t-1)\}/Ts \quad (7)$$

in which Ts is a time constant. This sub-model was first developed in the 1970s as part of the Bedford-Ouse model (Young, 2001), and is also used in the IHACRES model (Jakeman et al., 1990). The parameters of one of these two non-linear sub-models, when combined with those of the linear component, can describe the essential dynamics of the hydrological response of a catchment, and may be related to physical attributes (Ye et al., 1995; Hansen et al. 1996; Post et al., 1996; Schreider et al., 1997; Evans and Jakeman, 1998; Ye et al., 1998; Post and Jakeman, 1999).

A single-input, single-output first-order transfer function for rainfall-runoff modelling:

$$q(t) = \frac{b}{1 - a\, z^{-1}} u(t-d) \quad (8)$$

relates the runoff output at the time t, *q(t)*, to the rainfall input *u* in a model in which *a* is the recession or lag parameter, *b* is the system production or gain parameter, z^{-1} is the backward shift operator, and *d* is the time delay to the initial response (Young, 1998). The production or gain parameter (*b*) defines the production of runoff relative to rainfall, with a value of 1 implying that 100% of the rainfall generates runoff. The recession parameter (*a*) is a measure of the 'flashiness' of the rainfall-runoff behaviour; less flashy runoff is associated with values close to –1. This term is, however, normally expressed as a time-constant (TC) or a 'residence time' of water in the catchment runoff system. Higher TC values indicate longer residence times, or less flashy (more damped) rainfall-runoff behaviour. As the time constant can be equated to the residence time, it is also consistent with the time of concentration of the unit hydrograph, which is the time required for rain falling in the catchment to flow to the outlet. The time-constant is derived from the recession parameter, *a*, as in:

$$TC = \frac{t_{base}}{\log_e(a)} \qquad (9)$$

Here, t_{base} is the time-base of the sampling. The steady-state gain, ssG, which is the steady-state production relative to input, is derived from:

$$ssG = \frac{b}{1-a} \qquad (10)$$

A second-order model is required to model rainfall-runoff processes in catchments where there appears to be a dual response in the form of a rapidly-responding quickflow (or "fast flow"), and a delayed response ("slow flow"). In this case, a second-order model can be written in transfer function form as follows:

$$q(t) = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} u(t-d) \qquad (11)$$

where *q(t)* is the river flow at the time index *t*, a_i and b_i are the filter coefficients, z^{-1} is the backward shift operator, *d* is the time delay to the initial response, and *u* is the rainfall input.

This second-order model can be characterised by two real eigenvalues, and then factorized into first-order processes as follows:

$$q(t) = \left(\frac{b_s}{1-a_s z^{-1}}\right) u(t-d) + \left(\frac{b_q}{1-a_q z^{-1}}\right) u(t-d) \qquad (12)$$

Here a_s and b_s are respectively the recession and gain parameters of the slow flow, and a_q and b_q are the recession and gain parameter of the quick flow.

The optimal model DBM fit is defined by reference to two criteria: the coefficient of determination (R_t^2), and Young's information criterion (YIC), which helps to identify the most parsimonious (parametrically efficient) model.

$$R_t^2 = 1 - \frac{\hat{\sigma}^2}{\sigma_y^2} \qquad (13)$$

where

$$\sigma_y^2 = \frac{1}{N}\sum_{t=1}^{t=N}\left[y_t - \bar{y}\right]^2$$

$$\bar{y} = \frac{1}{N}\sum_{t=1}^{t=N} y_t$$

and

$$YIC = \log_e \frac{\hat{\sigma}^2}{\sigma_y^2} + \log_e\{NEVN\} \qquad (14)$$

where

$$NEVN = \frac{1}{np}\sum_{i=1}^{i=np}\frac{\hat{\sigma}^2 . \hat{p}_{ii}}{\hat{a}_i^2}$$

Here, $\hat{\sigma}^2$ is the variance of the model residuals, σ_y^2 is the variance of $[y_t - \bar{y}]$, *np = n +m+1* is the number of estimated parameters in the $\hat{\mathbf{a}}_N$ vector, $\hat{p}_{ii}$ is the *i*th diagonal element of the $\hat{\mathbf{P}}_t$ covariance matrix obtained from the estimation analysis and, $\hat{a}_i^2$ is the square of the *i*th parameter estimate in the $\hat{\mathbf{a}}_N$ vector.

YIC combines the residual variance term and a measure of parameter uncertainty; it is a logarithmic measure, which should be small (even negative). Since this measure includes the number of model parameters, it allows the user to avoid over-parameterised models by putting a penalty on them. However, this criterion should be used with caution, and in conjunction with R_t^2, since it may favour the wrong model. R_t^2 and YIC indicate how well the model can explain the data, and how well the parameters of the model are estimated (Young et al., 1997).

Application

The DBM methodology outlined above has been used to develop a strategy for modelling the rainfall-runoff relationship in the Mae Chaem catchment, initially using runoff data from the P14 gauging station about 10 km upstream from its junction with the River Ping. This has involved a series of modelling experiments as described in the sections below. The first step was to find an optimal model structure by fitting DBM models to successive annual records. Here, the objective was to examine whether a first- or second-order model was most appropriate, and the form of non-linear transformation required to provide an effective rainfall input series. This was initially undertaken with a rainfall record from a single raingauge, and subsequently evaluated further using different representations of the rainfall input. The model parameters were then examined for stability between years and rainfall representations, and were considered in terms of their physical implications. An attempt was then made to examine the use of DBM models in a semi-distributed manner, with a view to optimising the quality of simulations of the flow at the Huai Phung, Kong Kan and P14 gauging stations.

Model Results

The optimal model structure. As noted in the Introduction, previous DBM modelling for the Mae Chaem catchment has used a first-order model, but there is evidence that this may not be the optimal model, both given parameter variability between years, and time constants that conflict with other evidence of rapid response. A systematic comparative evaluation of the first- and second-order model structure reveals that a second-order model is indeed more generally applicable to the catchment. Table 1 summarises the results of fitting both models: (1,1,0) first-order and (2,2,1) second-order models were consistently the best structures in each case, as judged by the two criteria of the coefficient of determination (R_t^2), and Young's information criterion (YIC). Simultaneously, the suitability of the bi-linear power and storage transformations to provide effective rainfall were tested. In the Table, it can be seen that the (2,2,1) model structure tends to give a better R_t^2 than the (1,1,0) model, although the YIC criterion tends not to be as diagnostic (presumably because of the additional parameters requiring estimation for the second-order model). However, Figure 4 shows that the first-order

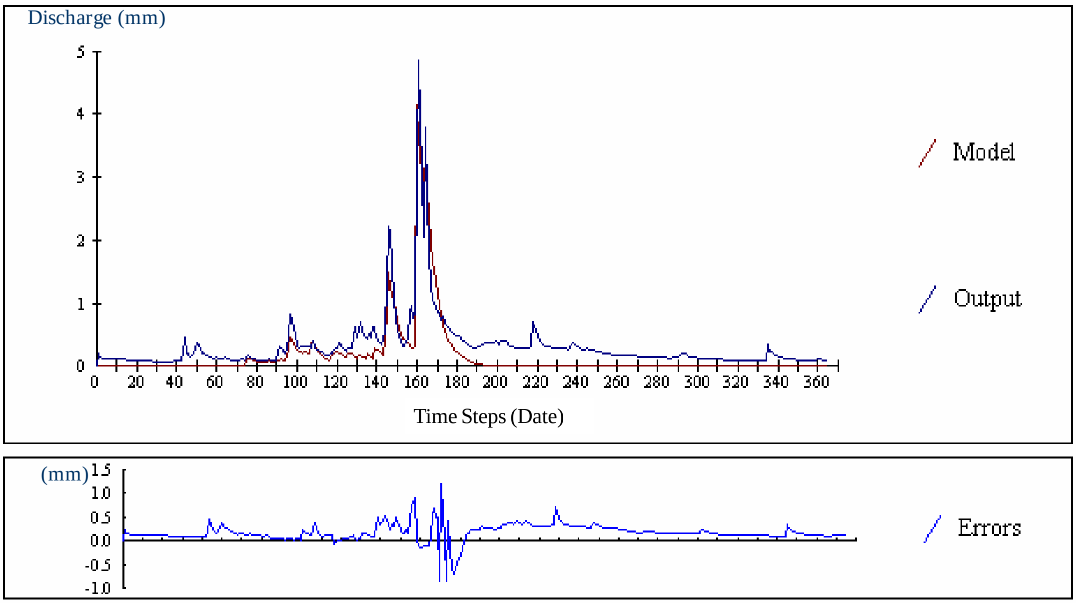


Figure 4 The first order [1,1,0] transfer function model using the bilinear power transformation, for the 1998 water year at the P14 river gauging station on the Mae Chaem River.

model, while it may capture the timing of the peak monsoon floods quite well, fits the late monsoon recession poorly. The preferred second-order model is that which employs the bi-linear power transformation; both test criteria are better in most years (except 1998). The implications of fitting the (2,2,1) model are firstly, that this implies that the rainfall-runoff behaviour is best represented by two storages with different response times; and secondly, that there is an advective delay (*d*) of one day in the hydrograph translation, which is consistent with the evidence comparing hydrograph peak times at the town of Mae Chaem and at the P14 gauge.

Variation in the quality of rainfall data. The quality of rainfall data is a constraint on successful rainfall-runoff modelling, and the existence of the GAME-T raingauge network provides an opportunity to test this effect. However, the GAME-T data are only available consistently for the period 2000-2002. In these three years it has been possible to fit the DBM transfer-function model to rainfall-runoff data for the catchment tributary to the P14 gauging station, but with the rainfall input (a) based on data from a single RID raingauge within the Mae Chaem catchment (the results of this being presented in Table 1); (b) based on data from the GAME-T network of 15 rain gauges (as in Figure 3), but with a simple average daily rainfall for these sites; and (c) based on the GAME-T data, with both Thiessen polygon area-weighting, and an altitude correction applied to reflect the increase in rainfall of 0.5mm per 100m above 1000m during the monsoon season (Lim, 2006).

When comparing (2,2,1) models based on respectively a single rain gauge, averaged rain gauge values, and area-weighted and altitude-corrected rainfall, it is clear that the relationship between rainfall from a single rain gauge and runoff provides the lowest R_t^2, while the averaged rainfall and flow give a better fit. However, area-weighted and altitude-corrected rainfall offers marginally the best model for predicting the Mae Chaem flow at P14, with the highest R_t^2 value (from 0.84-0.92 with the bi-linear power transformation). Thus, the better the spatial distribution of rain gauge data, the better the model fits the observed flow. The results for the more detailed rainfall data are summarised in Table 2, and Figure 5 shows the observed flow data and fitted model using the area- and altitude-corrected rainfall for 2002.

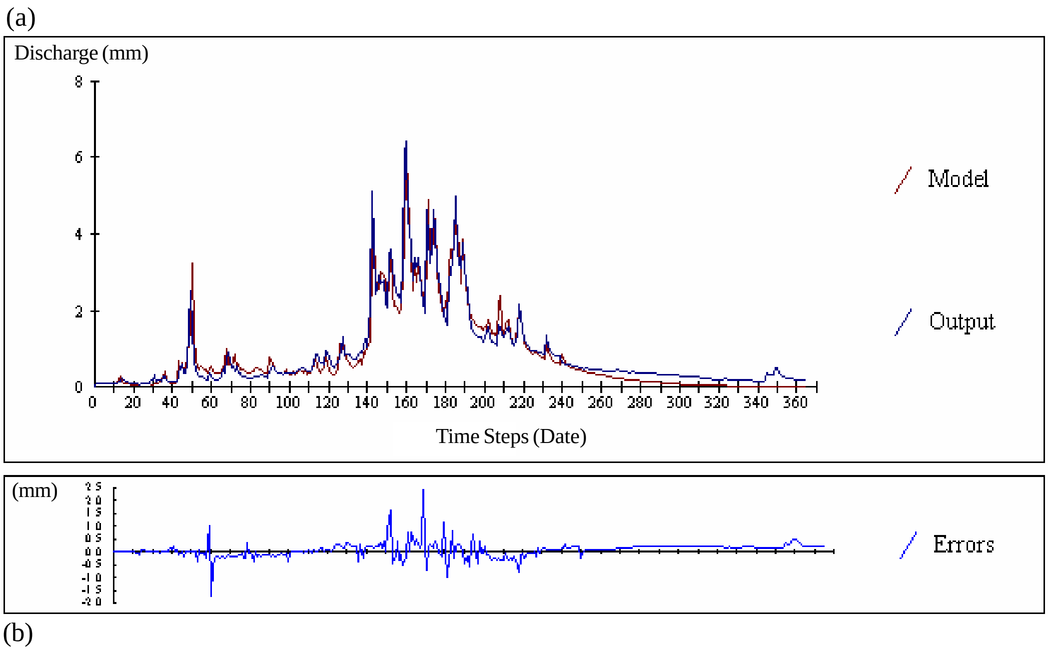


Figure 5 (a) The observed flow data at P14 on the Mae Chaem and the fitted model, using the area-corrected and altitude-adjusted rainfall for 2002 (the model is of order [2,2,1] with the bilinear model transformation, R_t^2 = 0.9210 and YIC = -7.2869). (b) The model errors.

Table 2 A comparison of the fitted (2,2,1) second-order models, using both the bi-linear power and the storage transformation, and applied to average rainfall input date, and area-weighted and altitude-corrected rainfall data for the period from 2000-2002.

Year	Average rainfall				Area-weighted and altitude-corrected rainfall			
	[2,2,1] bi-linear		[2,2,1] storage		[2,2,1] bi-linear		[2,2,1] storage	
	R_t^2	YIC	R_t^2	YIC	R_t^2	YIC	R_t^2	YIC
2000	0.8612	-7.2014	0.8122	-7.4081	0.8391	-6.4740	0.7683	-6.9800
2001	0.8512	-6.1324	0.7555	-3.1660	0.8631	-8.1736	0.8412	-6.0940
2002	0.9123	-6.2919	0.8610	-7.1851	0.9210	-7.2869	0.9039	-8.0364

The stability of parameter values. Since successive models fitted to runoff data relating to different monsoons are nevertheless models for the same catchment, it is interesting to consider the stability from year to year of the parameter values in the factorized version of the second-order model, in order to establish whether they can be regarded as representations of specific catchment properties. The second-order model has four parameters (the recession and gain parameters for each of the slow and fast flow components), and Table 3 summarises their fitted values for the (2,2,1) second-order models fitted, using the bi-linear power transformation, to the 1998-2002 single rain gauge data, the 2000-2002 average rainfall, and the 2000-2002 area- and altitude-corrected data.

Several tentative conclusions may be drawn from Table 3. Generally, stability of parameter values is most evident for the slow-flow coefficients (which might be expected to reflect more directly certain stable catchment properties), and especially for the gain coefficient b_s. The equivalent fast flow coefficients might be expected to vary more with the character of individual monsoons. There is a possibility, given these results, that continued use of the GAME-T rain gauge network could permit the definition of a parameter set for the Mae Chaem catchment that could be used for future forecasting. As a first approximation, this set might be a_s = 0.967, b_s = 0.003, a_q = 0.332 and b_q = 0.037 (averages of the values for the 2001-2002 simulations using the area-weighted and altitude-corrected rainfall). However, further data might refine

Table 3 The recession (a) and gain (b) parameters for second-order DBM models factorized into additive first-order processes for the slow flow (subscript ‘s’) and quick flow (subscript ‘q’) components (cf equation 12).

Rainfall/Year		a_s	b_s	a_q	b_q
Single	1998	0.9798	0.0014	0.6086	0.0176
	1999	0.9915	0.0010	0.8936	0.0105
	2000	0.9831	0.0025	0.5152	0.0170
	2001	0.9842	0.0022	0.3686	0.0304
	2002	0.9608	0.0078	0.1684	0.0215
Average	2000	0.9805	0.0021	0.4676	0.0293
	2001	0.9538	0.0031	0.1989	0.0409
	2002	0.9518	0.0036	0.2756	0.0349
Corrected	2000	0.9825	0.0025	0.5193	0.0291
	2001	0.9619	0.0036	0.2170	0.0442
	2002	0.9689	0.0031	0.3155	0.0438

these values, and the fast flow parameters might prove to vary according to properties of the monsoon regime, or the occurrence during the monsoon of cyclones, and possibly in due course, in predictable ways.

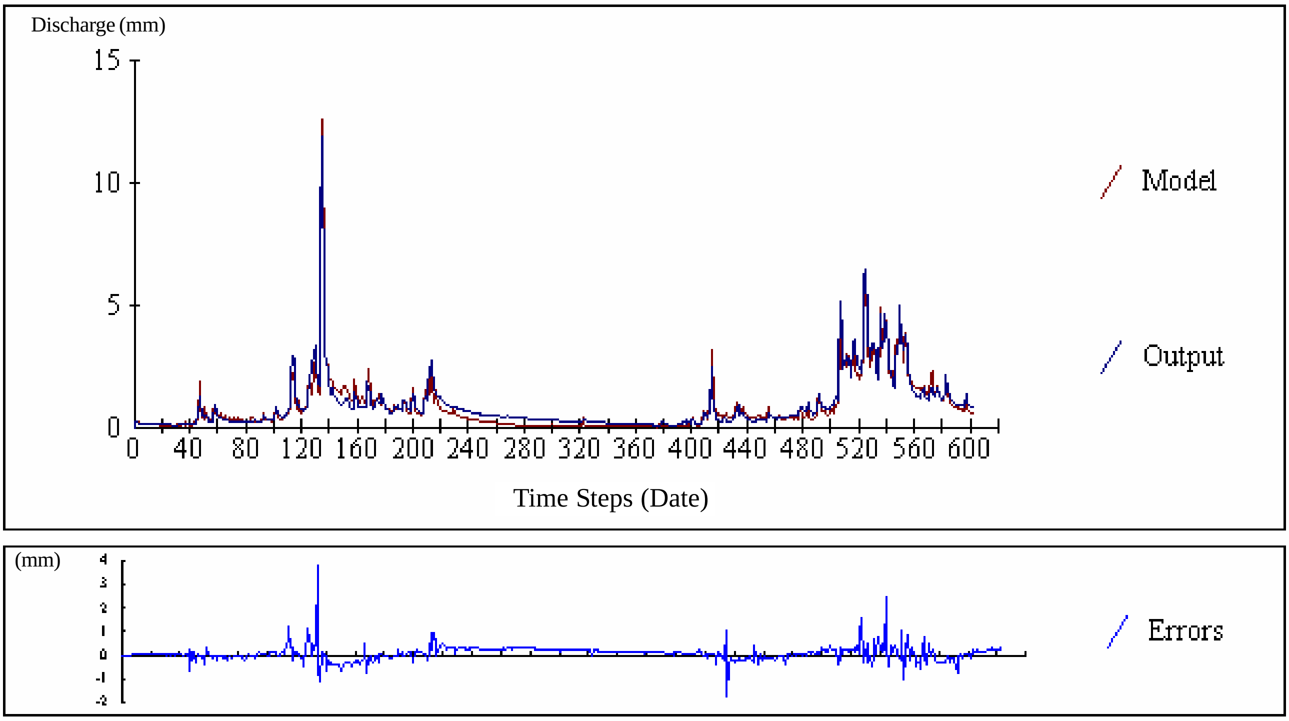
One way of testing the hypothesis that the DBM parameter values are stable enough from year to year to be used in forecasting is to calibrate parameter values on one set of data and test, or validate, on another. An experiment was therefore performed using the area-weighted and altitude-corrected rainfall data for two monsoon seasons (from 1 April 2001 to 25 November 2002; the first of these monsoons includes the 12 August 2001 flood) in order to define a DBM model. The parameters of this model were then used to simulate the flow in the April 2000-March 2001 period. The calibrated (2,2,1) model with the bi-linear power transformation had an R_t^2 criterion of 0.904, and when it was used with an unchanged parameter set to simulate the test period, the R_t^2 for this period was 0.736. This compares with the R_t^2 of 0.839 in Table 2 for the direct DBM model of the 2000 monsoon. The results of this test are shown in Figure 6. When the alternative storage transformation was used, both statistics indicated poorer fits, confirming the conclusion that the bi-linear power transformation is more suited to the rainfall-runoff behaviour of this catchment. A final test of transferability of the parameters was to use the average parameter set quoted above to simulate the flow in the period from 1 April 2001 to 25 November 2002. The result is a lower R_t^2 of 0.676. This probably reflects the fact that the fast flow gain parameter required in 2001 is 0.0442 (see Table 3), suggesting that fitting the unusually high peak flow of this year is sensitive to this parameter value, and implying that to undertake forecasting or continuous simulation will require refinement of the fast flow parameters so that they can reflect the probability and characteristics of extreme rainfall conditions.

Physical inferences from the parameter values. Second-order (2,2,1) models are interpreted physically through a parallel flow decomposition (Young et al., 1997; Ye et al., 1998) that allows them to be restated as the sum of two first-order processes representing slow and quick flow (equation 12). The data in Table 4 confirm that improving the quality of the rainfall data results in greater consistency in the flow parameters. The time constant for slow flow (TC_s) ranges from 25 to 118 days using single rain gauge inputs, but is 25-57 days using the best quality rainfall data. Using

Table 4 The time constant (TC), steady state gain (Gain), and percentage of flow associated with each of the slow flow (subscript 's') and fast flow (subscript 'q') components for the second-order models fitted to different rainfall data sets for the Mae Chaem catchment.

Rainfall/Year		TC_s	$Gain_s$	$Percent_s$	TC_q	$Gain_q$	$Percent_q$
Single	1998	49.04	0.0007	6.0	2.01	0.0109	94.0
	1999	117.77	0.0005	8.7	8.89	0.0055	91.3
	2000	58.62	0.0013	10.1	1.51	0.0112	89.9
	2001	62.63	0.0011	4.9	1.00	0.0222	95.1
	2002	25.01	0.0040	1.8	0.56	0.0184	82.3
Average	2000	50.69	0.0011	5.1	1.32	0.0200	94.9
	2001	21.15	0.0016	4.5	0.62	0.0341	95.5
	2002	20.25	0.0019	6.3	0.78	0.0273	93.7
Corrected	2000	56.73	0.0012	6.1	1.53	0.0192	93.9
	2001	25.77	0.0018	4.8	0.65	0.0363	95.2
	2002	31.63	0.0016	4.5	0.87	0.0333	95.5

(a)



(b)

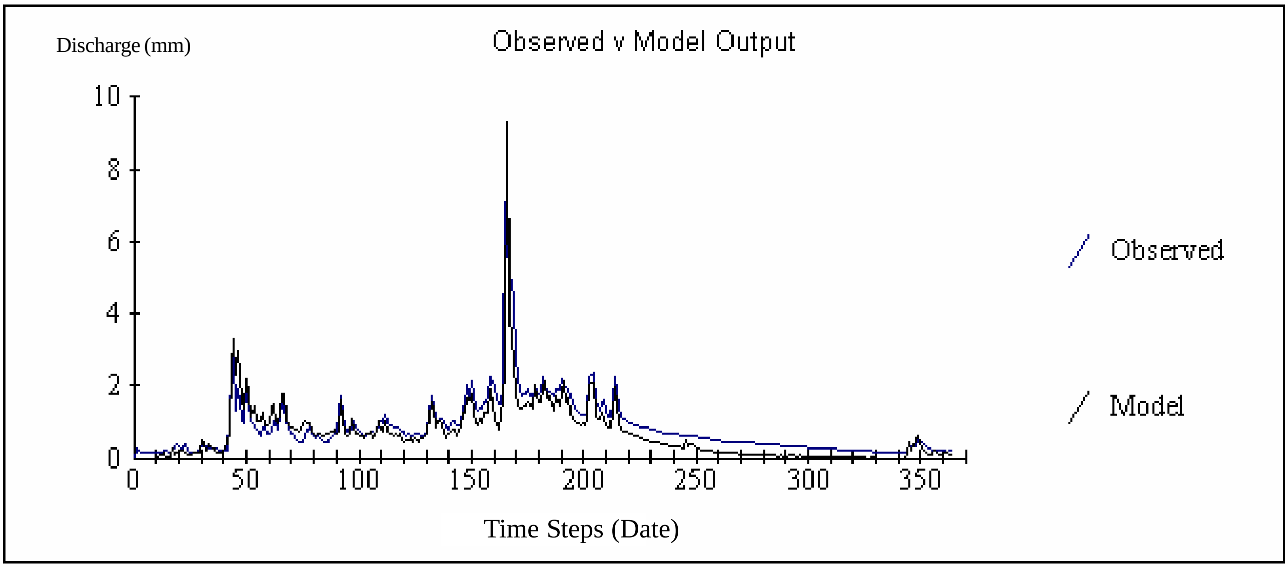


Figure 6 (a) The model calibration of the area-corrected and altitude-adjusted rainfall-runoff relationship (with bilinear model transformation) for the P14 catchment for the period from 1 April 2001 - 25 November 2002. (b) The errors of the model in (a). (c) A validation experiment using the [2,2,1] model from (a) to fit the data for April 2000 – March 2001; the transfer function parameters and rainfall filter were all as fitted in the calibration run whose results are shown in (a).

the area-weighted and altitude-corrected rainfall, the proportions of fast and slow flow (as defined in the context of this type and structure of model) are 95% and 5% respectively, the fast flow time constant (TC_q) is about 1 day, and the steady state gain for slow flow ($Gain_s$) is an order of magnitude lower than fast flow. In a catchment of this size, it may seem surprising that fast flow accounts for such a high proportion of runoff, since the factors that determine the fast flow might be expected to be related to network geometry and travel time effects, rather than to the pressure wave effects driving rapid subsurface flow response (Kirkby, 1976). However, the time step for the model is a daily one, and it includes a one-day advective delay which matches the observation that hydrographs tracked from Amphoe Mae Chaem (at approximately the catchment centroid) require a travel time of about one day to reach the downstream P14 gauging station. Thus, the additional time constant of one day for fast flow is effectively measuring the transformation of rainfall to runoff on what are the rather steep hillslopes of the catchment.

The table suggests that there are some consistent general properties of the rainfall-runoff system of the Mae Chaem catchment. However, as Young and Beven (1991) have shown, the potential for recovering the physical meaning of parameters is limited, since rather similar models such as the DBM model and IHACRES (Jakeman et al., 1990) result in quite different values of parameters such as time constants and percentages of runoff. This implies that it may be possible to draw inferences from variations in given parameters when a single model is applied to different catchments, but impossible to compare the numerical values of similar parameters obtained when fitting different models to the same catchment. We consider this further in a companion paper on the Mae Chaem rainfall-runoff data (Lim et al., 2007). There are some additional limitations to interpretation of the parameters. The fast pathway has a time constant that in two years is 0.65-0.87 of a day, which is less than data interval. This limits the reliability of the model, and suggests that higher resolution data may be needed to model this catchment satisfactorily. Also, the small proportion (4.5-6.3%) of water flowing along slower and presumably deeper groundwater pathways within the catchment may lead to uncertainty in the estimates of the time constants (Young et al., 1993). However, the time constants for water following the slow flow pathway are broadly comparable to the 54-141 day values for the Nymboida (1660 km^2), Timbarra (1720 km^2), Boyd (2670 km^2) and Clarence (4550 km^2) catchments in northern New South Wales, Australia, modelled by Hansen et al. (1996).

The potential for a semi-distributed use of DBM modelling.

Although the quality of rainfall data for the Mae Chaem catchment has been maximised by using Theissen polygon area-weighting and altitude adjustment, it remains the case that spatial variation in rainfall-runoff behaviour within this large catchment may reduce the quality of runoff simulation. This section examines whether simulation can be improved by disaggregating the catchment, and exploiting the existence of additional runoff data. There are two other river gauging stations on the main stream of the Mae Chaem, at Huai Phung and Kong Kan (the locations are shown in Figure 3), where the catchment areas are, respectively, 1,203.1 and 2,157.8 km^2. Table 5 shows the results of modelling the rainfall-runoff relationship for the catchments draining to the three gauging stations, for the period 1 April 2001 to 25 November 2002 (for the P14 results, see Figure 6). A high degree of similarity in model structure, test criteria and parameter values exists, although the most successful model fit appears to be for the entire catchment draining to the P14 gauge. This may reflect the relationship between model fit and data resolution. The P14 model has an advective delay parameter of one day, which is consistent with evidence of a hydrograph travel time from the mid-catchment town of Mae Chaem to P14 of one day. There is zero delay for the models fitted to the other two stations, which may simply mean that with daily data, no suitable

Table 5 DBM model parameters for the rainfall- runoff relationships for three gauging stations in the Mae Chaem catchment, fitted to data from 1 April 2001 to 25 November 2002.

Model Parameters	Huai Phung	Kong Kan	P14
Model structure	(2,2,0)	(2,2,0)	(2,2,1)
R_t^2	0.8062	0.8015	0.9040
YIC	-7.9708	-7.4488	-7.6429
TC_s	38.65	53.17	26.53
$Gain_s$	0.001	0.0007	0.0019
$Percent_s$	5.7	2.9	5.1
TC_q	1.75	1.26	0.70
$Gain_q$	0.0163	0.0254	0.0356
$Percent_q$	94.3	97.1	94.9

(< 24 hours) delay parameter can be defined. The lower R_t^2 for the upstream catchments may reflect the data resolution not being good enough to allow a model to be fitted that can measure the advective delay of the hydrograph within these catchments.

Discussion and Conclusion

The results outlined above suggest that data-based mechanistic models offer some potential for rainfall-runoff modelling in monsoon-dominated flow regimes in catchments of 1-10 x 10^3 km^2. They suggest that a second-order model which represents the catchment behaviour as a linear combination of a slow- and a quick-response component is a better model than the first-order model originally proposed for this catchment (Vongtanaboon and Chappell, 2004). There is evidence that the parameters of the slow-flow component are fairly stable from one monsoon season to the next, as would seem appropriate if they reflect the more stable underlying characteristics of the catchment. Of course, there are slow changes in catchment properties as a result of land use variations, and any attempt to define the slow-flow parameter values as a function of catchment properties would need to assess this possibility. The less stable quick-flow parameters may in part vary as they do because the small percentage of quick flow results in greater variance in parameter estimation. However, this could also reflect the transient character of individual monsoons, and again with longer data records, it may be possible to show systematic variations in these parameters with the timing of monsoon onset, the distribution of rainfall through the monsoon, the proportion of rainfall associated with cyclone activity, and even the amount of rainfall in the preceding dry season.

The (2,2,1) model structure which is appropriate for the P14 gauging station allows for an advective delay which is consistent with observational evidence. The DBM models fitted to the runoff data for the upstream gauges at Huai Phung and Kong Kan do not include an advective delay, being (2,2,0) models. This may reflect the data resolution (daily data), and the resulting inability to define a delay of less than one day. The consequence of this is that the lack of a delay parameter necessitates that some of the variance in the runoff data which reflects hydrograph transmission must be accounted for elsewhere in the model structure, and in other parameters. This may explain in part that the models for the upstream runoff data are slightly less good fits to the data than is the case for the downstream P14 station. The desirable data resolution thus reflects catchment scale; in the Mae Chaem, it appears that a resolution better than daily data may be required in sub-catchments up to about 2,000 km^2 in order to capture the advective character of the hydrograph, and that accordingly, the delay for P14 would then be of multiple time-steps. In the absence of the capacity to improve the resolution, however, it is necessary to accept that models may fit less well when the ratio of the data resolution to the observed time-of-travel of the hydrograph is greater than unity.

The evidence that better model fits are obtained with rainfall data which better reflect the spatial variability of rainfall input suggests that it would be worth exploring whether estimates of daily rainfall from the Om Koi rainfall radar which overlooks the Mai Chaem catchment could add further realism to the input data, in the form of continuously mapped distributions. This is an S-band Doppler 2.7-2.9 GHz radar which detects rain on the basis of intensity and velocity, with a maximum detection range of 480 km. The Om Koi data include an archive of radar data for a 200 km radius, with 5-minute resolution during the months of February to October; archived pre-monsoon rainfall data from an experimental raingauge network; and daily radiosonde data with temperature, wind, and aerosol data for the air column at Chiang Mai. These data have been used to explore aspects of the time-space variability of the rainfall in the region (Okumura et al., 2003). However, the lower technology runoff data collection maintained by the Royal Irrigation Department (using a network of stage readers) itself provides a basis for a simple modelling approach based on an enhancement of the traditional method based on upstream-downstream water level correlations by coupling this with incremental rainfall additions, as outlined above. In catchments bigger than about 1,000 km^2, flow modelling based only on rainfall-runoff relationships is always likely to be approximate and subject to error because of the difficulty of acquiring rainfall data representative of spatial variability, and the opportunity to use intermediate runoff data as a means of correcting the forecast downstream flow is a useful means of improving the quality of simulation. However, it would also be possible to use DBM rainfall-runoff models for the headwater catchment, and to combine simulated runoff at an intermediate gauging station with rainfall and runoff data for downstream sub-catchments.

Acknowledgements

The authors acknowledge of Hydrology and Fluid Dynamics Group, Environmental Science, Lancaster University for TFM Freeware and the assistance of Kosit Lorsirirat, Thada Sukhapunnaphan and Kanokporn Boochabun of the Royal Irrigation Department, Thailand, who have helped to provide the data and have contribued to the discussion of ideas.

References

Evans, J. P. and Jakeman, A. J. (1998) Development of a simple, catchment-scale, rainfall-evapotranspiration-runoff model. *Environmental Modelling & Software* 13: 385-393.

Hansen, D. P., Ye, W., Jakeman, A. J., Cooke, R., and Sharma, P. (1996) Analysis of the effect of rainfall and streamflow data quality and catchment dynamics on streamflow prediction using the rainfall-runoff model IHACRES. *Environmental Software* 11: 193-202.

Jakeman, A. J., Littlewood, I. G., and Whitehead, P.G. (1990) Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology* 117: 275-300.

Kirkby, M. J. (1976) Tests of the random network model, and its application to basin hydrology. *Earth Surface Processes* 1: 197-212.

Koichiro, K., Punyatrong, K., and Suzuki, M. (2001) Altitudial increase in rainfall in Mae Chaem watershed, Thailand. *Journal of the Meteorological Society of Japan* 1B. [Online URL:http://forester.uf.a.u-tokyo.ac.jp/~kuraji/MaeChaem/] accessed on April 1, 2006.

Lancaster University. (1996) *Catchment rainfall-runoff modelling using Transfer Function Model with nonlinear rainfall filtering, version 1.1 for Windows.*

Lim, H. S. (2006) *Rainfall-runoff behaviour of catchments in the seasonally humid tropics: North Thailand*. Unpublished PhD Thesis. Cambridge University, Cambridge.

Lim, H. S., Vongtanaboon, S., and Richards, K. S. (2007) Investigating the rainfall-runoff response of a small forested mountain catchment in the seasonally humid tropics using a modelling approach. *IUGG XXIV General Assembly "Earth: our changing planet"*. Perugia.

Okumura, K., Satomura, T., Oki, T., and Khantiyanan, W. (2003) Diurnal variation of precipitation by moving mesoscale systems: Radar observations in northern Thailand, *Geophysical Research Letters* 30(20): 2073.

Post, D. A. and Jakeman, A. J. (1999) Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling* 123: 91-104.

Post, D. A., Jakeman, A. J., Littlewood, I. G., Whitehead, P. G., and Jayasuriya, M. D. A. (1996) Modelling land-cover-induced variations in hydrologic response: Picaninny Creek, Victoria. *Ecological Modelling* 86: 177-182.

Schreider, S. Yu., Whetton, P. H., Jakeman, A. J., and Pittock, A. B. (1997) Runoff modelling for snow-affected catchments in the Australian alpine region, eastern Victoria. *Journal of Hydrology* 200: 1-23.

Vongtanaboon, S. (2004) *Parsimonious modelling the rainfall-runoff behaviour of large catchments in Thailand*. Unpublished PhD Thesis. Lancaster University, Lancaster.

Vongtanaboon, S. and Chappell, N. A. (2004) DBM rainfall-runoff modelling of large rainforest catchments in Thailand. In *Forests and Water in Warm, Humid Asia* (Sidle, R. C., Tani, M., Nik, A. R., and Tadese, T. A., eds.), pp.252–255. Disaster Prevention Research Institute, Uiji.

Ye, W., Jakeman, A. J., and Barnes, C. J. (1995) A parametrically efficient model for prediction of streamflow in an Australian benchmark catchment with complex storage dynamics. *Environment International* 21: 539-544.

Ye W., Jakeman, A. J., and Young, P. C. (1998) Identification of improved rainfall-runoff models for an ephemeral low-yielding Australian catchment. *Environmental Modelling & Software* 13: 59-74.

Young, P. C. (1998) Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling & Software* 13: 105-122.

Young, P. C. (2001) Data-based mechanistic modelling and validation of rainfall-flow processes. In *Model validation: perspectives in hydrological science* (Anderson, M. G. and Bates, P. D, eds.), pp. 117-161. Chichester, J.Wiley.

Young, P. C. and Beven, K. J. (1994) Data-based mechanistic modelling and the rainfall-flow non-linearity. *Environmetrics* 5: 335-363.

Young, P. C. and Beven, K. (1991) Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments – Comment. *Journal of Hydrology* 129: 389-396.

Young, P. C., Chotai, A., and Tych, W. (1993) Identification, Estimation and True Digital Control of Glasshouse Systems. In *The computerized Greenhouse: Automatic Control Application in Plant Production* (Hashimoto, Y. et al., eds.), pp. 3-50. Academic Press, New York.

Young, P. C., Jakeman, A. J., and Post, D. A. (1997) Recent advances in the data-based modelling and analysis of hydrological systems. *Water Scence and Technology* 36: 99-116.