



HAL
open science

The PDM rainfall-runoff model

R. J. Moore

► **To cite this version:**

R. J. Moore. The PDM rainfall-runoff model. Hydrology and Earth System Sciences Discussions, 2007, 11 (1), pp.483-499. hal-00305633

HAL Id: hal-00305633

<https://hal.science/hal-00305633>

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The PDM rainfall-runoff model

R.J. Moore

Centre for Ecology and Hydrology, Wallingford, Oxon, OX10 8BB, UK

Email: rm@ceh.ac.uk

Abstract

The Probability Distributed Model, or PDM, has evolved as a toolkit of model functions that together constitute a lumped rainfall–runoff model capable of representing a variety of catchment-scale hydrological behaviours. Runoff production is represented as a saturation excess runoff process controlled by the absorption capacity (of the canopy, surface and soil) whose variability within the catchment is characterised by a probability density function of chosen form. Soil drainage to groundwater is controlled by the water content in excess of a tension threshold, optionally inhibited by the water content of the receiving groundwater store. Alternatively, a proportional split of runoff to fast (surface storage) and slow (groundwater) pathways can be invoked with no explicit soil drainage function. Recursive solutions to the Horton-Izzard equation are provided for routing flows through these pathways, conveniently considered to yield the surface runoff and baseflow components of the total flow. An alternative routing function employs a transfer function that is discretely-coincident to a cascade of two linear reservoirs in series. For real-time flow forecasting applications, the PDM is complemented by updating methods based on error prediction and state-correction approaches. The PDM has been widely applied throughout the world, both for operational and design purposes. This experience has allowed the PDM to evolve to its current form as a practical toolkit for rainfall-runoff modelling and forecasting.

Keywords: rainfall-runoff model, PDM, flooding, updating, forecasting

Introduction

The Probability Distributed Model or PDM is a fairly general conceptual rainfall–runoff model which transforms rainfall and potential evaporation data to flow at the catchment outlet (Moore, 1985, 1999; Moore and Bell, 2002; Moore *et al.*, 2005; Institute of Hydrology, 1992, 1996; CEH Wallingford, 2005). Figure 1 illustrates the general form of the model. Runoff production at a point in the catchment is controlled by the absorption capacity of the soil (treated together with canopy and surface detention) to take up water. This can be conceptualised as a simple store with a given storage capacity. By considering that different points in a catchment have differing storage capacities and that the spatial variation of capacity can be described by a probability distribution, it is possible to formulate a simple runoff production model which integrates the point runoffs to yield the catchment surface runoff into surface storage. Groundwater recharge from the soil moisture store passes into subsurface storage. The outflow from surface and subsurface storages, together with any fixed flow representing, say, compensation releases

from reservoirs or constant abstractions, forms the model output.

The components of the PDM model are described in detail as the main subject of this paper. Formulations for the probability-distributed soil moisture store are first derived based on mass balance principles with the addition of rainfall, losses to evaporation, drainage to groundwater (recharge) and the production of direct runoff. The direct runoff and recharge are routed via surface and subsurface storages, representing fast and slow pathways to the basin outlet. Representations of these surface and subsurface storages — by a choice of Horton-Izzard equation solution or by a transfer function discretely coincident with a cascade of two linear reservoirs — are set down next. This is followed by an overview of the model parameters and their calibration, discussion of example applications and a summary of the typical model form that is invoked within the PDM toolkit. Having completed the description of the PDM as a simulation model, methods of updating provided for real-time forecasting applications are outlined. Methods

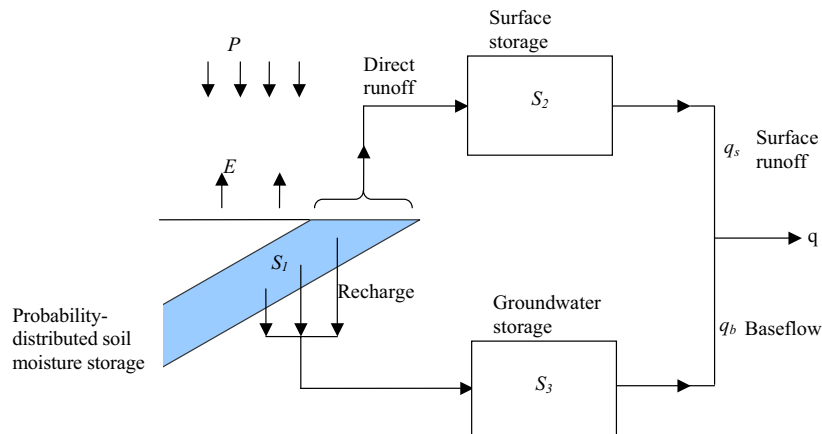


Fig. 1. The PDM rainfall-runoff model

of empirical state-correction for the PDM and more generic error-prediction techniques are described, and their relative merits discussed. The paper ends with a historical perspective on the evolution of the PDM to its current state of development together with some concluding remarks.

Probability-distributed soil moisture store

Consider that runoff production at any point within a river basin may be conceptualised as a single storage, or tank, of capacity c' , representing the absorption capacity of the soil column at that point. The storage takes up water from rainfall, P , and loses water by evaporation, E , until either the storage fills and spills, generating direct runoff, q , or empties and ceases to lose water by evaporation. Figure 2(a) depicts such a storage, whose behaviour may be expressed mathematically by

$$q' = \begin{cases} P - E - (c' - S_0) & P > c' + E \\ 0 & P \leq c' + E \end{cases} \quad (1)$$

where S_0 is the initial depth of water in storage, and where P , E and q' represent the depth of rainfall, evaporation and the resulting direct runoff over the interval being considered. Now consider that runoff production at every point within a river basin may be similarly described, each point differing from another only with regard to the storage capacity. The storage capacity at any point, c , may then be considered as a random variate with probability density function, $f(c)$, so that the proportion of the river basin with depths in the range $(c, c + dc)$ will be $f(c)dc$.

The water balance for a river basin assumed to have storage capacities distributed in this way may be constructed

as follows. First imagine that stores of all possible different depths are arranged in order of depth and with their open tops arranged at the same height: this results in a wedge-shaped diagram as depicted in Fig. 2(b). If the basin is initially dry so that all stores are empty and rain falls at a net rate P for a unit duration, then stores will fill to a depth P unless they are of lesser depth than P when they will fill and spill. During the interval the shallowest stores will start generating direct runoff and at the end of the interval stores of depth P will just begin to produce runoff, so that the upper triangular area in Fig. 2(c) denotes the depth produced from stores of different depth over the unit interval. Since, in general, there are more stores of one depth than another the actual runoff produced over the basin must be obtained by weighting the depth produced by a store of a given depth by its frequency of occurrence, as expressed by $f(c)$. Now, at the end of the interval stores of depth less than P are generating runoff: let this critical capacity below which all stores are full at some time t be denoted by $C^* \equiv C^*(t)$ ($C^* = P$ in the present example). The proportion of the basin containing stores of capacity less than or equal to C^* is

$$\text{prob}(c \leq C^*) = F(C^*) = \int_0^{C^*} f(c)dc. \quad (2)$$

The function $F(\cdot)$ is the distribution function of store capacity and is related to the density function, $f(c)$, through the relation $f(c) = dF(c)/dc$. This proportion is also the proportion of the basin generating runoff, so that the contributing area at time t for a basin of area A is

$$A_c(t) = F(C^*(t))A. \quad (3)$$

The instantaneous direct runoff rate per unit area from the basin is the product of the net rainfall rate, $\pi(t)$, and the

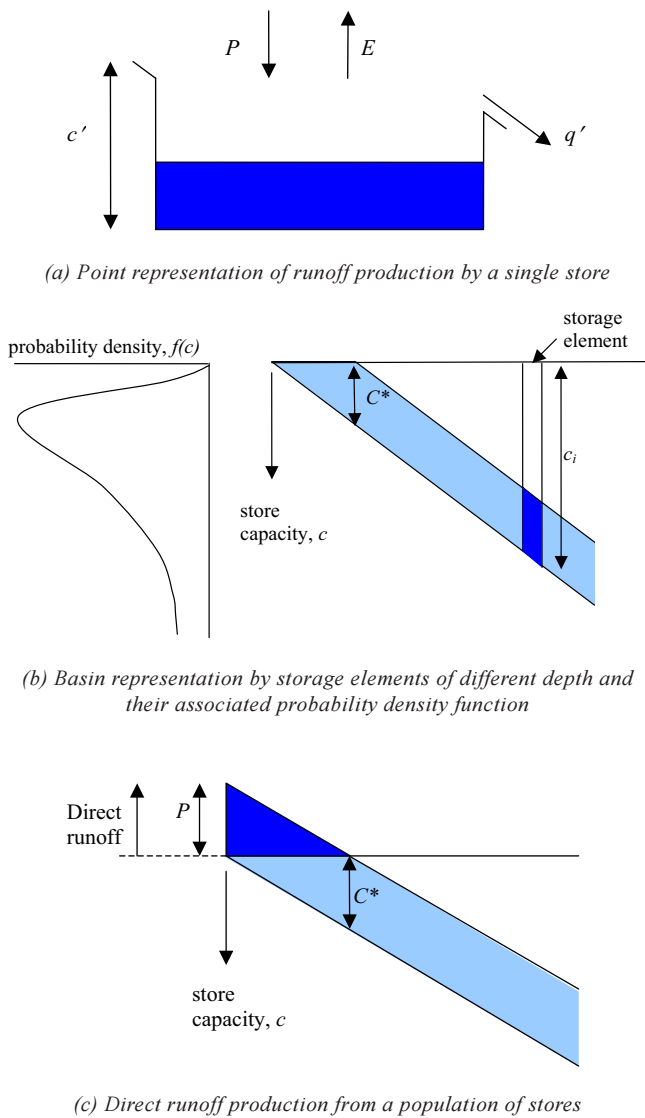


Fig. 2. Definition diagrams for the probability-distributed interacting storage capacity component

proportion of the basin generating runoff, $F(C^*(t))$; that is

$$q(t) = \pi(t) F(C^*(t)). \quad (4)$$

During the i 'th wet interval, $(t, t + \Delta t)$, suppose rainfall and potential evaporation occur at constant rates P_i and E_i , so that net rainfall $\pi_i = P_i - E_i$. Then the critical capacity, $C^*(\tau)$, will increase over the interval according to

$$C^*(\tau) = C^*(t) + \pi_i (\tau - t) \quad t \leq \tau \leq t + \Delta t, \quad (5)$$

the contributing area will expand according to Eqn. (3), and the volume of basin direct runoff per unit area produced over this interval will be

$$V(t + \Delta t) = \int_t^{t + \Delta t} q(\tau) d\tau = \int_{C^*(t)}^{C^*(t + \Delta t)} F(c) dc. \quad (6)$$

During dry periods potential evaporation will deplete the water content of each storage. It will be assumed during such depletion periods that water moves between storages of different depths so as to equalise the depth of stored water at different points within the basin. Thus at any time all stores will have a water content, C^* , irrespective of their capacity, unless this is less than C^* when they will be full: the water level profile across stores of different depths will therefore always be of the simple form shown in Fig. 2(c). The assumption which allows redistribution of water between storages of different size during depletion periods is particularly important for real-time applications of the model where the possibility of updating the store contents is envisaged. Moore (1985) shows how this assumption, when not invoked, leads to a more complex water accounting procedure which is less amenable to real-time empirical state adjustment schemes. Particularly important is that a unique relationship exists between the water in storage over the basin as a whole, $S(t)$, and the critical capacity, $C^*(t)$, and in turn to the instantaneous rate of basin runoff production, $q(t)$. Specifically, and referring to Fig. 2(c), it is clear that the total water in storage over the basin is

$$\begin{aligned} S(t) &= \int_0^{C^*(t)} cf(c) dc + C^*(t) \int_{C^*(t)}^{\infty} f(c) dc \\ &= \int_0^{C^*(t)} (1 - F(c)) dc. \end{aligned} \quad (7)$$

For a given value of storage, $S(t)$, this can be used to obtain $C^*(t)$ which allows the volume of direct runoff, $V(t + \Delta t)$, to be calculated using Eqn. (6) together with Eqn. (5).

The dependence of evaporation loss on soil moisture content is introduced by assuming the following simple function between the ratio of actual to potential evaporation, E'_i / E_i , and soil moisture deficit, $S_{max} - S(t)$:

$$\frac{E'_i}{E_i} = 1 - \left\{ \frac{(S_{max} - S(t))}{S_{max}} \right\}^{b_e}; \quad (8)$$

either a linear ($b_e = 1$ so $E'_i = (S(t) / S_{max}) E_i$) or quadratic form ($b_e = 2$) is usually assumed. Here, S_{max} is the total available storage, and is given by

$$S_{max} = \int_0^{\infty} cf(c) dc = \int_0^{\infty} (1 - F(c)) dc = \bar{c}, \quad (9)$$

where \bar{c} is the mean storage capacity over the basin.

Further loss as recharge to groundwater may be introduced by assuming that the rate of drainage over the interval, d_i ,

depends linearly on basin soil moisture content at the start of the interval i.e.

$$d_i = k_g^{-1} (S(t) - S_i)^{b_g} \quad (10)$$

where k_g is the drainage time constant and b_g the exponent of the recharge function (usually set to 1) and S_i is the threshold storage below which there is no drainage, water being held under soil tension. An alternative formulation is available which allows recharge to depend on both soil and groundwater storage for use in catchments where soil/groundwater interactions are important. Consider recharge into a groundwater store of maximum capacity S_g^{\max} . Then a groundwater deficit ratio may be defined as

$$g(t) = \frac{S_g^{\max} - S_g(t)}{S_g^{\max}} \quad (11)$$

where $S_g(t)$ denotes the groundwater storage at time t . This ratio can be used to define a groundwater demand factor between 0 and 1:

$$f(t) = \begin{cases} \left(\frac{g(t)}{\alpha}\right)^\beta & g(t) < \alpha \\ 1 & \text{otherwise} \end{cases} \quad (12)$$

which achieves a maximum for values of the deficit ratio $g(t)$ in excess of α . It is then reasonable to suppose that the recharge depth over the interval, D_i , will increase with soil storage, $S(t)$, and with the groundwater demand factor, $f(t)$, according to

$$D_i = (D_{sat} + (S_{\max} - D_{sat})f(t)) \frac{S(t)}{S_{\max}} \quad (13)$$

Here the recharge depth at saturation $D_{sat} = q_{sat} \Delta t$, with q_{sat} the outflow from the groundwater storage when $S_g(t)$ equals S_g^{\max} (i.e. the maximum rate of recharge). Note that the drainage rate over the interval is $d_i = D_i / \Delta t$. There are thus only three parameters: α , β and q_{sat} (with S_g^{\max} thereby implied from its storage function). It is seen that, for a saturated soil store, recharge is diminished when the groundwater demand factor is less than α , when the soil ceases to be freely draining. This formulation derives from a reparameterised form of percolation model used in the National Weather Service rainfall-runoff model (Burnash *et al.*, 1973; Gupta and Sarooshian, 1983).

A third recharge formulation is available which assumes that there is no soil drainage, d_i . Direct runoff is split between a fraction α which goes to make up surface runoff and a fraction $(1-\alpha)$ going to groundwater storage.

With both losses to evaporation and recharge, the net rainfall, π_i , may be defined in general as

$$\pi_i = P_i - E_i' - d_i \quad (14)$$

During a period when no runoff generation occurs then, for this general case, soil moisture storage accounting simply involves the calculation

$$S(\tau) = S(t) + \pi_i(\tau - t) \quad t \leq \tau \leq t + \Delta t, 0 \leq S(\tau) \leq S_{\max} \quad (15)$$

When runoff generation does occur then the volume of runoff produced, $V(t+\Delta t)$, is obtained using Eqn. (6), and then continuity gives the replenished storage as

$$S(t + \Delta t) = \begin{cases} S(t) + \pi_i \Delta t - V(t + \Delta t) & S(t + \Delta t) \leq S_{\max} \\ S_{\max} & \text{otherwise} \end{cases} \quad (16)$$

If basin storage is fully replenished within the interval $(t, t+\Delta t)$ then $V(t+\Delta t)$ should be computed from continuity as

$$V(t + \Delta t) = \pi_i \Delta t - (S_{\max} - S(t)) \quad (17)$$

The above completes the procedure for soil moisture accounting and determining the value of runoff production according to a probability-distributed storage capacity model. Figure 3 provides a graphical representation of this procedure for a wet interval $(t, t+\Delta t)$ during which soil moisture storage is added to by an amount $\Delta S(t+\Delta t) = \pi_i \Delta t - V(t+\Delta t)$, and a volume of direct runoff, $V(t+\Delta t)$, is generated.

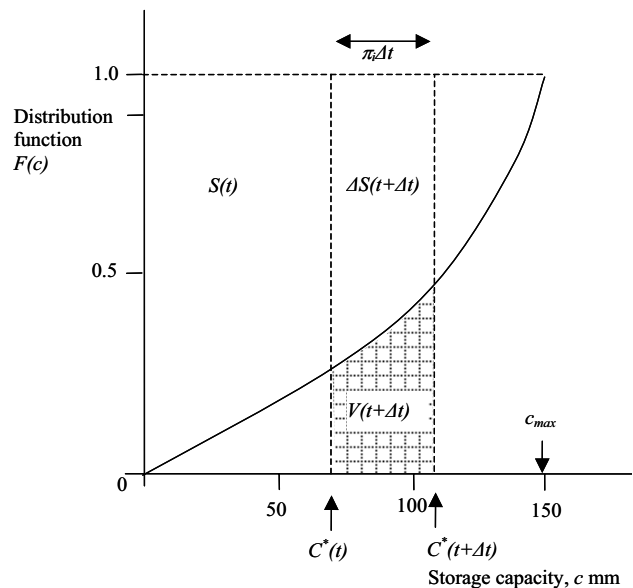


Fig. 3. The storage capacity distribution function used to calculate basin moisture storage, critical capacity, and direct runoff according to the probability-distributed interacting storage capacity model.

A specific application of the procedure can be developed for a given choice of probability density function. Analytical solutions of the integrals in the probability-distributed storage capacity model component (specifically Eqns. (6) and (7)) are presented in Appendix A for a range of possible distribution types. After a number of trials on alternative distributions, a Pareto distribution of storage capacity is now most widely used in practice and will be used here to illustrate application of the method. The distribution function and probability density function for this distribution are

$$F(c) = 1 - \left(1 - \frac{c}{c_{\max}}\right)^b \quad 0 \leq c \leq c_{\max} \quad (18)$$

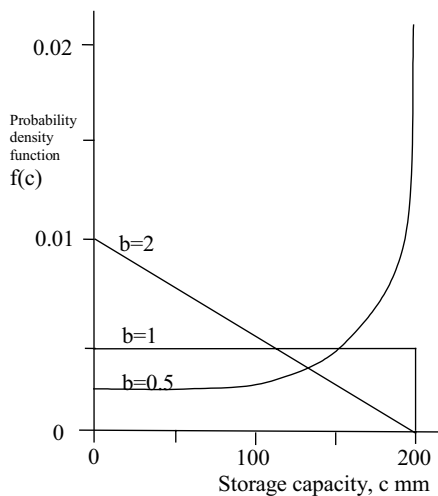
$$f(c) = \frac{dF(c)}{dc} = \frac{b}{c_{\max}} \left(1 - \frac{c}{c_{\max}}\right)^{b-1} \quad 0 \leq c \leq c_{\max} \quad (19)$$

where parameter c_{\max} is the maximum storage capacity in the basin, and parameter b controls the degree of spatial variability of storage capacity over the basin. These functions are illustrated in Fig. 4: note that the rectangular distribution is obtained as a special case when $b=1$, and $b=0$ implies a constant storage capacity over the entire basin. The following relations apply for Pareto distributed storage capacities:

$$S_{\max} = c_{\max} / (b + 1), \quad (20a)$$

$$S(t) = S_{\max} \left\{ 1 - \left(1 - C^*(t) / c_{\max}\right)^{b+1} \right\}, \quad (20b)$$

$$C^*(t) = c_{\max} \left\{ 1 - \left(1 - S(t) / S_{\max}\right)^{1/(b+1)} \right\}, \quad (20c)$$

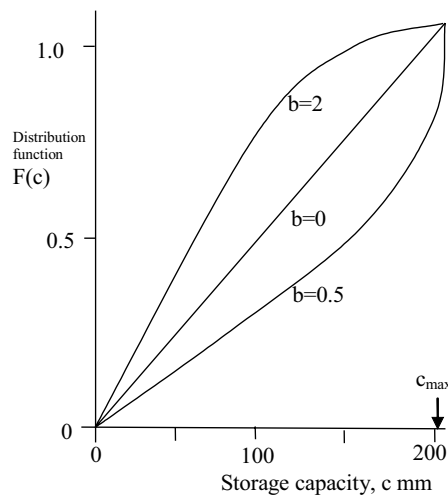


(a) Probability density function

$$V(t + \Delta t) = \pi_i \Delta t - S_{\max} \left\{ \left(1 - C^*(t) / c_{\max}\right)^{b+1} - \left(1 - C^*(t + \Delta t) / c_{\max}\right)^{b+1} \right\}. \quad (20d)$$

The relationship between rainfall and runoff implied by the above expressions, for given conditions of soil moisture, is presented in Fig. 5. A related, if not similar, procedure forms the basis of the Xinanjiang model developed by Ren Jun Zhao and co-workers in China (Zhao and Zhuang, 1963; Zhao *et al.*, 1980; Zhao and Liu, 1995). More recently it has been popularised and extended in the form of the Arno rainfall-runoff model in Italy (Todini, 1996) and, for large-scale applications, the VIC land surface model in the USA (Wood *et al.*, 1992). Indeed, Moore (1985) traces the origins of such probability-distributed principles in hydrology back to the pioneering contribution of Bagrov in 1950, working in what was then the USSR.

Extended forms of these relations are available in the PDM toolkit to represent the case where storage capacity is limited by both the upper bound c_{\max} and a lower bound c_{\min} such that $c_{\min} \leq c \leq c_{\max}$ in the Pareto distribution (see Appendix A and Moore and Bell, 2002). Note that when $C^*(t)$ drops to the value of c_{\min} as a result of evaporation, the basin water storage $S(t)$ also equals this value. Equation (15) is used to update $S(t)$ during ‘dry periods’ allowing it to fall below c_{\min} and ultimately to zero. Only once net rainfall has replenished water storage above c_{\min} does runoff generation occur and the calculation of $C^*(t)$ and $V(t)$ is resumed. Thus the initial catchment response to rainfall can be moderated or delayed if the minimum storage capacity is taken to be greater than zero.



(b) Distribution function

Fig. 4. The Pareto distribution of storage capacity.

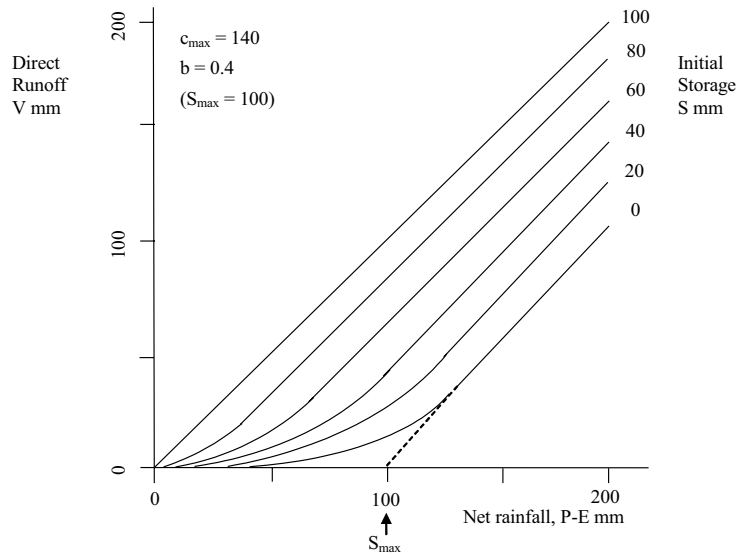


Fig. 5. Rainfall-runoff relationship for the probability-distributed interacting storage capacity model, using the Pareto distribution of storage capacity.

Surface and subsurface storages

The probability-distributed store model partitions rainfall into direct runoff, groundwater recharge and soil moisture storage. Direct runoff is routed through surface storage: a ‘fast response system’ representing channel and other fast translation flow paths. Groundwater recharge from soil water drainage is routed through subsurface storage: a ‘slow response system’ representing groundwater and other slow flow paths. Both routing systems can be defined in the PDM by a variety of non-linear storage reservoirs or by a cascade of two linear reservoirs (expressed as an equivalent second order transfer function model constrained to preserve continuity).

The nonlinear storage model is specified by the Horton-Izzard equation (Dooge, 1973)

$$\frac{dq}{dt} = a(u - q)^b, \quad q > 0, -\infty < b < 1, \quad (21)$$

where $q \equiv q(t)$ is the rate of outflow and $u \equiv u(t)$ is the rate of inflow to the store per unit area, and a and b are parameters. This derives from the continuity equation for the store

$$\frac{dS}{dt} = u - q, \quad (22)$$

and the nonlinear storage form of the momentum equation

$$q = k S^m, \quad k > 0, m > 0 \quad (23)$$

where $S \equiv S(t)$ is the volume of water held in the storage per unit area, k is the storage rate coefficient and m the store exponent. Note that $a = mk^{1/m}$ and $b = (m-1)/m$.

Recursive solutions of the Horton-Izzard equation are provided in the PDM for a choice of non-linear storage form: linear, quadratic, cubic, exponential and general non-linear. A cubic form is usually considered most appropriate to represent the groundwater storage. In this case where $q = k S^3$, an approximate solution utilising a method due to Smith (1977) yields the following recursive equation for storage, given a constant input u over the interval $(t, t + \Delta t)$:

$$S(t + \Delta t) = S(t) - \frac{1}{3kS^2(t)} \{ \exp(-3kS^2(t)\Delta t) - 1 \} (u - kS^3(t)). \quad (24)$$

Discharge may then be obtained simply using the nonlinear relation

$$q(t + \Delta t) = k S^3(t + \Delta t). \quad (25)$$

Solutions for the other nonlinear forms are presented in Moore (1983) and Appendix A of Moore and Bell (2002). When used to represent groundwater storage, the input u will be the drainage rate, d_s , from the soil moisture store and the output $q(t)$ will be the ‘baseflow’ component of flow $q_b(t)$. Explicit allowance for groundwater abstractions is incorporated in an extended version of the PDM which can also make use of well level data (Moore and Bell, 2002).

The most commonly used representation of the surface storage component is a cascade of two linear reservoirs,

with time constants k_1 and k_2 , expressed as the discretely coincident transfer function model (O'Connor, 1982):

$$q_t = -\delta_1 q_{t-1} - \delta_2 q_{t-2} + \omega_0 u_t + \omega_1 u_{t-1} \quad (26)$$

with

$$\delta_1 = -(\delta_1^* + \delta_2^*), \quad \delta_2 = \delta_1^* \delta_2^*, \quad \delta_1^* = \exp(-\Delta t/k_1), \quad \delta_2^* = \exp(-\Delta t/k_2)$$

$$\begin{aligned} \omega_0 &= \frac{k_1(\delta_1^* - 1) - k_2(\delta_2^* - 1)}{k_2 - k_1} & k_1 \neq k_2 \\ \omega_1 &= \frac{k_2(\delta_2^* - 1)\delta_1^* - k_1(\delta_1^* - 1)\delta_2^*}{k_2 - k_1} & k_1 \neq k_2 \\ \omega_0 &= 1 - (1 + \Delta t/k_1)\delta_1^* & k_1 = k_2 \\ \omega_1 &= (\delta_1^* - 1 + \Delta t/k_1)\delta_1^* & k_1 = k_2. \end{aligned} \quad (27)$$

Here Δt is the time interval between times $t-1$ and t and it is assumed that the input u_t is constant over this interval. In this case the input is the volume of direct runoff, $V(t)$, generated from the probability-distributed soil moisture store and the output q_t will be the surface flow component of the total basin runoff, $q_s(t)$. The total basin flow is given by $q_s(t) + q_b(t)$, plus a constant flow, q_c , representing any returns or abstractions.

Model parameters, calibration and example application

The parameter and structure options in the model are summarised in Table 1. Note that a rainfall factor, f_c , is incorporated in the model to allow conversion of a rainfall observation to rainfall, P , thereby compensating for effects such as lack of raingauge representativeness. The time constants k_g and k_b are equivalent to k^l in the general non-

Table 1. PDM model parameters

Parameter name	Unit	Description
f_c	none	rainfall factor
t_d	hour	time delay
Probability-distributed store		
c_{min}	mm	minimum store capacity
c_{max}	mm	maximum store capacity
b	none	exponent of Pareto distribution controlling spatial variability of store capacity
Evaporation function		
b_e	none	exponent in actual evaporation function
Recharge function		
1: Standard		
k_g	hour mm $b_g - 1$	groundwater recharge time constant
b_g	none	exponent of recharge function
S_t^g	mm	soil tension storage capacity
2: Demand-based		
α	none	groundwater deficit ratio threshold
β	none	exponent in groundwater demand factor function
q_{sat}	mm h ⁻¹	maximum rate of recharge
3: Splitting		
α	none	runoff factor controlling the split of rainfall to surface and groundwater storage routing when no soil recharge is allowed
Surface routing		
k_1, k_2	hour	time constants of cascade of two linear reservoirs
Groundwater storage routing		
k_b	hour mm ^{m-1}	baseflow time constant
m	none	exponent of baseflow non-linear storage
Constant flow addition		
q_c	m ³ s ⁻¹	constant flow representing returns/abstractions

linear storage function $q=kS^m$.

The calibration of the PDM model is carried out within a generic Model Calibration Shell environment. This Calibration Shell provides for both automatic optimisation and informal visually-interactive parameter estimation. The former uses a simplex direct search procedure (Nelder and Mead, 1965) modified following suggestions made by Gill *et al.* (1981). Informal estimation is supported by an interactive visualisation tool which allows the user to see the changing hydrograph response as a chosen parameter value is varied. This can be an invaluable aid to understanding the model response and the nature of the dependence between parameters. Error response function plots for a selected pair of parameters can also be used to investigate parameter interdependence.

The seasonal response of a model, dominated by aspects of the model structure and a certain subset of the model parameters controlling the water balance, can be investigated through a preliminary calibration at a daily time-step. A 15-minute time-step can then be used to establish the model parameters dominating the short-term dynamics of runoff response and translation. Storm events spanning several years can be included in the optimisation at a 15-minute time-step. This is achieved by prescribing the event periods to be included in the optimisation and switching to a daily time-step between events, for purposes of continuous water accounting. Other time-steps to the 15-minute and daily interval can be used. If long (several years) continuous 15-minute datasets are available, then modelling can be done at this interval. The Calibration Shell visualisation facilities

can be used to look at the response over seasons or years as well as zoomed-in to flood hydrographs of special interest. The objective function used to assess model performance can be censored to exclude flows below a minimum value so that it is not unduly influenced by long hydrograph recessions.

The PDM model has been widely applied to a variety of catchments in different countries. These include England, Wales, Scotland, Belgium, Hong Kong, India, South Java, Thailand and China. An example of its use as a simulation model is shown in Fig. 6. The flows are for Beverley Brook gauged at Wimbledon Common in London within the Thames basin, draining an area of 44 km². HYRAD recalibrated radar rainfall data (Moore, 1999), which combines radar and raingauge information, is used to form the catchment average rainfall employed as input to the model in this case.

Summary of the PDM rainfall-runoff model

Whilst the design of the PDM model provides for a range of different structural forms, the most commonly used configuration comprises the following components:

- (i) A *probability-distributed soil moisture storage component* to effect separation between direct runoff and subsurface runoff. This is based on a Pareto distribution of soil moisture storage capacity over the catchment.

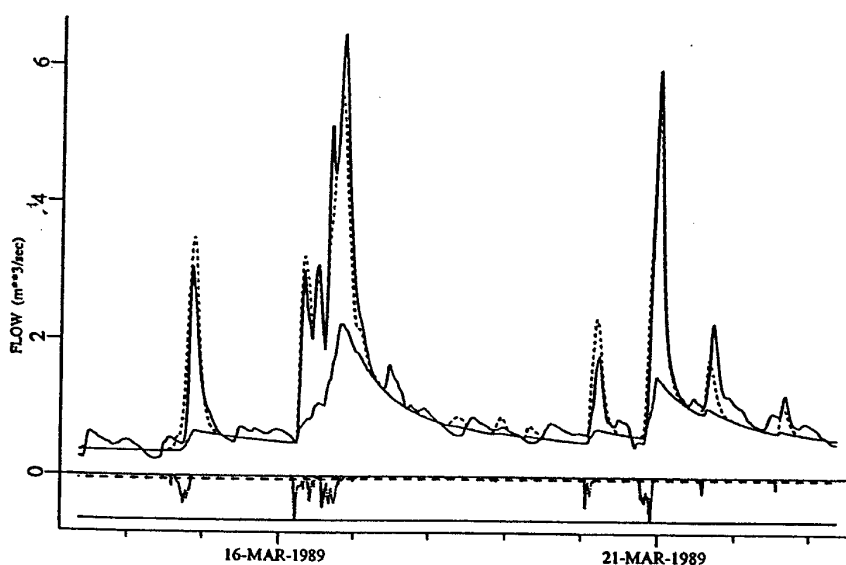


Fig. 6. PDM model simulation of Beverley Brook at Wimbledon Common using HYRAD recalibrated radar rainfall data as input. Observed flow: bold line; simulated flow: dashed line; "baseflow": small dots. The negative ordinate shows soil moisture deficit as a dashed line and rainfall on a proportional scale.

- (ii) A *surface storage component* which transforms direct runoff to surface runoff. This employs a two-linear reservoir cascade formulated as a transfer function with dependence on two past outputs and the current and previous input. The coincidence of this four parameter discrete time model with the two parameter continuous time model that preserves continuity allows the transfer function model to be (a) re-parameterised as a two parameter model, and (b) to be used at different discrete time intervals.
- (iii) A *groundwater storage component* which receives drainage water from the distributed soil moisture storage as input and contributes the groundwater component of total runoff as output. A cubic non-linear storage routing function is adopted to effect this transformation.

Methods are available to update the PDM model with reference to observed flows for real-time flow forecasting applications. An *empirical state-correction scheme* provides a range of options for correcting internal model water contents or flow rates to yield more accurate updated model forecasts. As an alternative, an *ARMA error-prediction scheme* is available which exploits the persistence in model errors to obtain improved forecasts. These methods are described in outline next.

Updating the PDM rainfall-runoff model

INTRODUCTION

If observed flows are not used, except for initialisation, a model is said to be operating in simulation mode, acting as a function which transforms rainfall and potential evaporation to river flow. A model which has been calibrated in simulation mode may be extended to use observed flows by addition of further structure and associated parameters. These might take the form of rules for adjusting model states (state-correction) or predicting future errors (error-prediction). The former are heavily dependent on the structure of the simulation mode model, whilst the latter are essentially independent. Parameter-adjustment is not considered in the PDM. The view is taken that this approach confuses the issue of correct model identification, which is properly carried out through a controlled calibration procedure. Parameter variability is better addressed by improving the structural form of the model than by tracking its variation in real-time.

A model incorporating observed flows either through state-correction or error-prediction will be said to be

operating in updating mode. These two updating methods are available for use with the PDM and are detailed below.

STATE CORRECTION

The term 'state' is used to describe a variable of a model which mediates between inputs to the model and the model output (Szollosi-Nagy, 1976). In the case of the PDM rainfall-runoff model, the main input is rainfall and basin flow is the model output. Typical state variables are the water contents of the surface and groundwater stores, S_2 and S_3 , and of the probability-distributed soil storage, S_1 (Fig. 1). The flow rates out of the conceptual stores can also be regarded as state variables: examples are q_s , the flow out of the surface storage, and q_b , the flow out of the groundwater storage.

When an error, $\varepsilon = Q - q = Q - (q_s + q_b)$, occurs between the model prediction, q , and the observed value of basin runoff, Q , it would seem sensible to 'attribute the blame' to mis-specification of the state variables and attempt to 'correct' the state values to achieve concordance between observed and model predicted flow. Mis-specification may, for example, have arisen through errors in rainfall measurement which, as a result of the model water accounting procedure, are manifested through the values of the store water contents, or equivalently the flow rates out of the stores. A formal approach to state correction is provided by the Kalman filter algorithm (Jazwinski, 1970; Gelb, 1974; Moore and Weiss, 1980a,b). This provides an optimal adjustment scheme for incorporating observations, through a set of linear operations, for linear dynamic systems subject to random variations which may not necessarily be Gaussian in form. For non-linear dynamic models, such as the PDM, an extended form of Kalman filter based on a linearisation approximation is required which is no longer optimal in the adjustment it provides. The implication of this is that simpler, intuitive adjustment schemes can be devised which potentially provide better adjustments than the more complex and formal extensions of the Kalman filter which accommodate non-linear dynamics through approximations. Such schemes which make physically sensible adjustments are called 'empirical state adjustment schemes'. A simple example is the apportioning of the error, ε , between the surface and groundwater stores of the PDM in proportion to their contribution to the total flow. Mathematically this may be expressed as

$$q_b^* = q_b + \alpha g_b \varepsilon \quad (28a)$$

$$q_s^* = q_s + (1 - \alpha) g_s \varepsilon \quad (28b)$$

where

$$\alpha = q_b / (q_s + q_b) \quad (29)$$

and the superscript * indicates the value after adjustment. The 'gain' coefficients, g_b and g_s , when equal to unity, yield the result that $q_b^* + q_s^*$ equals the observed flow, Q , thus achieving exact correction of the model flow to equal the observed value. Values of the coefficients other than unity allow for different adjustments to be made, and g_b and g_s can be regarded as model parameters whose values are established through optimisation to achieve the 'best' fit between state-adjusted forecasts and observed flows. A generalisation of the above is to define α to be

$$\alpha = \frac{q_b}{\beta_1 q_s + \beta_2 q_b} \quad (30)$$

and to choose the incidental parameters β_1 and β_2 to weight the apportionment towards or away from one of the flow components; in practice β_1 and β_2 are assigned values of 10 and 1.1 to apportion more of the error adjustment to the surface store. Note that the adjustment is carried out at every time step and the time subscripts have been omitted for notational simplicity. The scheme with α defined by Eqn. (29) is referred to as the proportional adjustment scheme and that defined by Eqn. (30) is the super-proportional adjustment scheme. Replacing α and $(1-\alpha)$ in Eqns. (28a,b) by unity yields the simplest non-proportional adjustment scheme.

An adjustment to the probability-distributed soil moisture, $S \equiv S_1$, may also be made, either of the proportional form

$$S^* = S + \alpha g_g \varepsilon \quad (28c)$$

or the direct form of gain with α equal to unity.

It should be noted that all the above forms of adjustment utilise the same basic form of adjustment employed by the Kalman filter in which an updated state estimate is formed from the sum of the current state value and the model error multiplied by a gain coefficient. However, instead of defining the gain statistically, as the ratio of the uncertainty in the observation to that of the current state value, it is first related to a physical apportionment rule multiplied by a gain factor. This gain factor acts as a relaxation coefficient which is estimated through an off-line optimisation using past flood event data.

State-correction is essentially a form of negative feedback and, although usually very effective, this feedback can sometimes give rise to an over- or under-shooting behaviour characterised by high accuracy at short lead times but with degraded accuracy at moderate lead times before a recovery

in accuracy at longer lead times. This behaviour appears to be associated with a combination of some or all of the following: large gain factors, time lags between the correction of a state value and the appearance of an effect on the modelled flow, and rapid increases in the model error (often due to timing errors on the rising limb). The latter is also a problem for error prediction schemes. Optimal values for the gain factors tend to be greater than unity (over-relaxation) whilst time lags can occur because correction of soil moisture may not affect runoff until the next wet period.

ERROR PREDICTION

State-correction techniques have been developed based on adjustment of the water content of conceptual storage elements in the belief that the main cause of the discrepancy between observed and modelled runoff will arise from errors in estimating basin average rainfall, which in turn accumulate as errors in water storage content. Rather than attribute the cause directly and devise empirical adjustment procedures the structure of the errors may be analysed and predictors of future errors developed based on this structure which can then be used to obtain improved flow forecasts. A feature of errors from a conceptual rainfall-runoff model is that there is a tendency for errors to persist so that sequences of positive errors (underestimation) or negative errors (overestimation) are common. This dependence structure in the error sequence may be exploited by developing error predictors which incorporate this structure and allow future errors to be predicted. Error-prediction is now a well established technique for forecast updating in real time (Box and Jenkins, 1970; Moore, 1982). Error-prediction is available as an alternative to empirical state correction in the PDM. Predictions of the error are added to the deterministic model prediction to obtain the updated model forecast of flows. In contrast to the state correction scheme, which adjusts values internally within the model, the error prediction scheme is wholly external to the deterministic model operation. The importance of this is that error prediction may be used in combination with any model, be it of TF, conceptual or 'physics-based' form, and for representing rainfall-runoff or channel flow processes.

Consider that $q_{t+\ell}$ is the forecast of the observed flow, $Q_{t+\ell}$, at some time $t+\ell$, made using the PDM rainfall-runoff model. Since $q_{t+\ell}$ will have essentially been obtained by transformation of rainfall into flow through the PDM model conceptualisation of the catchment, it will not have used previous observed values of flow, except for the purposes of model initialisation. It will consequently be referred to as a simulation-mode forecast to distinguish it from a real-time, updated forecast which incorporates information from

observed flows.

The error, $\eta_{t+\ell}$, associated with this simulation-mode forecast is defined through the relation

$$Q_{t+\ell} = q_{t+\ell} + \eta_{t+\ell}. \quad (31)$$

If the simulation-mode error $\eta_{t+\ell}$ may be predicted using an error predictor which exploits the dependence structure of these errors, then an improved forecast may be obtained.

Let $\eta_{t+\ell|t}$ denote a prediction of the simulation-mode error, $\eta_{t+\ell}$, made ℓ steps ahead from a forecast origin at time t using an error predictor. (The suffix notation $t+\ell|t$ should be read as a forecast at time $t+\ell$ given information up to time t .) Then a real-time forecast, $q_{t+\ell|t}$, made ℓ time units ahead from a forecast origin at time t may be expressed as follows:

$$q_{t+\ell|t} = q_{t+\ell} + \eta_{t+\ell|t}. \quad (32)$$

The real-time forecast error is

$$a_{t+\ell|t} = Q_{t+\ell} - q_{t+\ell|t} \quad (33)$$

which, depending on the performance of the error predictor, should be smaller than the simulation-mode forecast error

$$\eta_{t+\ell} = Q_{t+\ell} - q_{t+\ell}. \quad (34)$$

Turning now to an appropriate form of error predictor it is clear that a structure which incorporates dependence on past simulation-mode errors is required. Thus the autoregressive (AR) model

$$\eta_t = -\phi_1 \eta_{t-1} - \phi_2 \eta_{t-2} - \dots - \phi_z \eta_{t-z} + a_t \quad (35)$$

is an obvious candidate, where a_t is the residual error (uncorrelated), and $\{\phi_i\}$ are parameters. However, a more parsimonious form of model is of the autoregressive-moving average (ARMA) form

$$\eta_t = -\phi_1 \eta_{t-1} - \phi_2 \eta_{t-2} - \dots - \phi_p \eta_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t \quad (36)$$

which incorporates dependence of past residual errors, a_{t-1} , a_{t-2} , ...

In general, the number of parameters $p+q$ associated with the ARMA model will be less than the number z associated with the AR model, in order to achieve as good a level of approximation to the true simulation-mode error structure. The ARMA model may be used to give the following error predictor:

$$\eta_{t+\ell|t} = -\phi_1 \eta_{t+\ell-1|t} - \phi_2 \eta_{t+\ell-2|t} - \dots - \phi_p \eta_{t+\ell-p|t} + \theta_1 a_{t+\ell-1|t} + \theta_2 a_{t+\ell-2|t} + \dots + \theta_q a_{t+\ell-q|t}, \quad \ell = 1, 2, \dots \quad (37)$$

where

$$a_{t+\ell-i|t} = \begin{cases} 0 & \ell - i > 0 \\ a_{t+\ell-i} & \text{otherwise} \end{cases} \quad (38)$$

and $a_{t+\ell-i}$ is the one-step ahead prediction error

$$\begin{aligned} a_{t+\ell-i} &\equiv a_{t+\ell-i|t+\ell-i-1} = \eta_{t+\ell-i} - \eta_{t+\ell-i|t+\ell-i-1} \\ &= Q_{t+\ell-i} - q_{t+\ell-i|t+\ell-i-1}, \end{aligned} \quad (39)$$

and

$$\eta_{t+\ell-i|t} = \eta_{t+\ell-i} = Q_{t+\ell-i} - q_{t+\ell-i} \quad \text{for } \ell - i \leq 0. \quad (40)$$

The prediction Eqn. (37) is used recursively to produce the error predictions $\eta_{t+1|t}$, $\eta_{t+2|t}$, ..., $\eta_{t+\ell|t}$, from the available values of a_t , a_{t-1} , ... and η_t , η_{t-1} , ...

Using this error predictor methodology, the PDM model simulation-mode forecasts, $q_{t+j|t}$, may be updated using the error prediction $\eta_{t+\ell|t}$, obtained from Eqn. (37) (and the related Eqns. (38–40)), to calculate the required real-time forecast, $q_{t+\ell|t}$, according to Eqn. (32). Note that this real-time forecast incorporates information from the most recent observations of flow through the error predictor, and specifically through calculation of the one-step ahead forecast errors, $a_{t+\ell-i}$, according to Eqn. (39).

Alternative error predictor schemes may be devised by working with other definitions of the basic errors: for example by using proportional errors. One such scheme can be formulated by starting with the logarithmic model

$$\log Q_{t+\ell} = \log q_{t+\ell} + \eta_{t+\ell} \quad (41)$$

so that the simulation-mode error is now defined as

$$\eta_{t+\ell} = \log(Q_{t+\ell}/q_{t+\ell}). \quad (42)$$

An error predictor for $\eta_{t+\ell}$ may be formulated in the normal way using Eqns. (37) and (38) with the one-step ahead prediction error given by

$$a_{t+\ell-i} = \eta_{t+\ell-i} - \eta_{t+\ell-i|t+\ell-i-1}. \quad (40)$$

Instead of Eqn. (32) the real-time forecast, $q_{t+\ell|t}$, takes the form

$$q_{t+\ell|t} = q_{t+\ell} \exp(\eta_{t+\ell|t}). \quad (41)$$

The PDM uses automatic optimisation to estimate the ARMA

error-predictor parameters for a chosen model structure. Often a third order autoregressive, with dependence on three past model errors, provides an appropriate choice for UK conditions and a 15 minute model/data time interval.

DISCUSSION OF UPDATING THE PDM

Whilst error-prediction provides a general technique which is easy to apply, its performance in providing improved forecasts will depend on the degree of persistence in the model errors. Unfortunately, in the vicinity of the rising limb and peak of the flood hydrograph this persistence is least and errors show a tendency to oscillate rapidly and most widely; dependence is at its strongest for errors on the falling limb, where improved forecast performance matters least. In addition, timing errors in the model forecast may lead to erroneous error predictions being made, a problem which is also shared by the technique of state-correction. The general applicability and popularity of error-prediction as an updating tool commends its use as an ‘off-the-shelf’ technique, but empirical state adjustment schemes should also be considered as viable alternatives to the use of error-prediction. State-correction is normally the preferred choice with the PDM model.

Historical perspective and concluding remarks

The PDM rainfall-runoff model had its origins in the search for model formulations that were well suited to automatic parameter estimation: specifically, models that avoided threshold behaviours and were parsimonious of parameters. Initially, this focused on invoking a probability distribution of storage capacity to replace a simple single store representation of runoff production (Moore and Clarke, 1981). The threshold behaviour of the single store results in a discontinuity in the gradient of the model’s objective function and difficulties of parameter optimisation. These problems were circumvented by the probability-distributed representation and, more importantly, led to a more realistic representation of runoff production across a catchment. A simplification of the theory (Moore, 1985) — allowing storage elements to interact with each other so as to equalise the depth of stored water across a catchment — led to the formulations now employed in the PDM. Over time, solutions were obtained for a range of distributions including exponential, Pareto (reflected power), rectangular, triangular, power and lognormal (see Appendix A). The theory also allowed for representations of evaporation and drainage to recharge as a function of catchment soil moisture.

These early developments employed convolution to route direct runoff (and drainage to recharge) to the basin outlet, using the probability-distributed principle applied to the time-of-travel to the catchment outlet. Specifically, the probability density function (pdf) of translation time was taken as equivalent to the kernel function or instantaneous unit hydrograph. Invoking a pdf of inverse Gaussian form provided a physical link with the St. Venant equations of open-channel flow and the convection-diffusion equation. It was not until 1986 that the model schematic of Fig. 1 first appeared (see Fig. 3 of Moore, 1986). The parallel routing formulation made explicit by this schematic was inspired by the simplified catchment model schematic of Dooge (Fig. 1-8 in Dooge, 1973; Fig. 2.3 in Dooge and O’Kane, 2003). Here, soil moisture feedback controls the partitioning of the direct storm and groundwater response of total runoff, deliberately avoiding the less identifiable separation of overland flow and interflow. Convolution was replaced by non-linear storage routing as being simpler and easier to update for real-time application. The exposition of the Horton-Izzard equation in Dooge (1973), and reference to the source and related texts, provided the foundation for the recursive solutions to this equation used for storage routing in the PDM (Moore, 1983); Moore and Bell (2002) provides further details and reference sources. This was supplemented by the transfer function representation of a cascade of two linear reservoirs efficiently parameterised via one or two storage coefficients (O’Connor, 1982) as a further routing option.

This early development of a generalised rainfall-model toolkit was undertaken deliberating without recourse to a ‘brand name’ association, recognising that many models share rather similar conceptual elements. However, with the need to develop the River Flow Forecasting System (RFFS) for operational use, the requirement for a software product arose and the PDM brand name introduced (Moore and Jones, 1991; Institute of Hydrology, 1992).

As the PDM toolkit evolved to represent a wider range of hydrological behaviours it became clear that the original aim of a model well suited to automatic parameter estimation had been compromised. It became evident that a rainfall-runoff model that performed well was typically not easy to optimise automatically; the parallel configuration of routing stores and more complex recharge functions particularly introduced parameter interdependence. Thus a more realistic representation of hydrological behaviour is likely to have problems of parameter estimation. An acceptance of this position has led to increasing reliance being placed on manual calibration using interactive visual support tools, using automatic optimisation only for late-stage refinement. It is not clear that recent advances in calibration of rainfall-

runoff models (Duan *et al.*, 2003) will change this position radically, although the ability to search the parameter space more comprehensively with increased computer power may help.

Updating of the PDM for real-time forecasting applications benefited from experience gained in the 1970s and 1980s (Moore and Weiss, 1982a,b; Moore, 1982, 1986). This experience concerned the recursive state-parameter estimation techniques based on the Kalman filter (Jazwinski, 1970) and the transfer function noise models of ARMA form popularised by Box and Jenkins (1970). Formal Kalman filter techniques for state-correction were abandoned in favour of the simpler empirical state-correction methods described here. The ARMA noise models were used as the basis of the error-predictor technique available as an alternative method of updating the PDM forecast. The parameters of these updating methods are well suited to automatic optimisation, once the PDM simulation model parameters have been first estimated.

In spite of its widespread application, the PDM has not been formally published in extensive form in the open literature. The detail has been confined to the manuals and guides provided with the software product (CEH Wallingford, 2005). Only rather general reviews of the PDM have been given in published papers on flood forecasting techniques (Moore, 1999; Moore *et al.*, 2005). This paper has rectified this omission.

References

- Box, G.E.P. and Jenkins, G.M., 1970. *Time series analysis forecasting and control*, Holden Day, San Francisco, 553pp.
- Burnash, R.J.C., Ferral, R.L. and McGuire, R.A., 1973. *A generalized streamflow simulation system: conceptual modelling for digital computers*. Report of the Joint Federal State River Forecast Centre, U.S. National Weather Service and California Department of Water Resources, Sacramento.
- CEH Wallingford, 2005. *PDM Rainfall-Runoff Model*. Version 2.2, Centre for Ecology & Hydrology, Wallingford. (Includes Guide, Practical User Guides, User Manual and Training Exercises).
- Dooge, J.C.I., 1973. *Linear theory of hydrologic systems*. Tech. Bull. 1468, Agric. Res. Service, US Dept. Agric., Washington, 327 pp.
- Dooge, J.C.I. and O'Kane, J.P., 2003. *Deterministic methods in systems hydrology*. Balkema, The Netherlands. 309pp.
- Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.) 2003. *Calibration of watershed models*. Water Science and Application 6, American Geophysical Union, Washington, USA. 345pp.
- Gelb, A. ed., 1974. *Applied optimal estimation*. MIT Press, Cambridge, USA. 374 pp.
- Gill, P.E., Murray, W. and Wright, M.H., 1981. *Practical optimisation*. Academic Press, London, 401pp.
- Gupta, V.K. and Sorooshian, S., 1983. Uniqueness and observability of conceptual rainfall-runoff model parameters: The Percolation process examined. *Water Resour. Res.*, **19**, 269–276.
- Institute of Hydrology, 1992. *PDM: A generalized rainfall-runoff model for real-time use, Developers' Training Course*. National Rivers Authority River Flow Forecasting System, Version 1.0, March 1992, 26pp.
- Institute of Hydrology, 1996. *A guide to the PDM*. Version 1.0. Wallingford, UK. 45pp.
- Jazwinski, A.H., 1970. *Stochastic processes and filtering theory*. Academic Press, New York, USA. 376 pp.
- Moore, R.J., 1982. Transfer functions, noise predictors and the forecasting of flood events in real-time. In: *Statistical analysis of rainfall and runoff*, V.P. Singh (Ed.), Water Resources Publ., Littleton, Colorado, USA. 229–250.
- Moore, R.J., 1983. *Flood forecasting techniques*. WMO/UNDP Regional Training Seminar on Flood Forecasting, Bangkok, Thailand. 37pp.
- Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.*, **30**, 273–297.
- Moore, R.J., 1986. Advances in real-time flood forecasting practice. *Symposium on Flood Warning Systems*, Winter meeting of the River Engineering Section, Inst. Water Engineers and Scientists, 23pp.
- Moore, R.J., 1999. Real-time flood forecasting systems: Perspectives and prospects. In: *Floods and landslides: Integrated Risk Assessment*, R. Casale and C. Margottini (Eds.), Springer, Berlin, Germany. Chapter 11, 147–189.
- Moore, R.J. and Bell, V.A., 2002. Incorporation of groundwater losses and well level data in rainfall-runoff models illustrated using the PDM. *Hydro. Earth System Sci.*, **6**, 25–38.
- Moore, R.J., Bell, V.A. and Jones, D.A., 2005. Forecasting for flood warning. *C. R. Geoscience*, **337**, 203–217.
- Moore, R.J. and Clarke, R.T., 1981. A distribution function approach to rainfall-runoff modelling. *Water Resour. Res.*, **17**, 1367–1382.
- Moore, R.J. and Jones, D.A., 1991. A river flow forecasting system for region-wide application. Invited paper, *MAFF Conference of River and Coastal Engineers 1991*, 8–10 July 1991, Loughborough University, 12pp.
- Moore, R.J. and Weiss, G., 1980a. Recursive parameter estimation of a non-linear flow forecasting model using the extended Kalman filter. In: *Real-time hydrological forecasting and control*, P.E. O'Connell, (Ed.), *Proc. 1st Int. Workshop*, July 1977, Institute of Hydrology, Wallingford, UK. 264pp.
- Moore, R.J. and Weiss, G., 1980b. Real-time parameter estimation of a nonlinear catchment model using extended Kalman filters. In: *Real-time forecasting/control of water resource systems*, E.G. Wood and A. Szollosi-Nagy (Eds.), Pergamon Press, Oxford. 83–92.
- Nelder, J.A. and Mead, R., 1965. A simplex method for function minimisation. *Computer J.*, **7**, 308–313.
- O'Connor, K.M., 1982. Derivation of discretely coincident forms of continuous linear time-invariant models using the transfer function approach. *J. Hydrol.*, **59**, 1–48.
- Smith, J.M., 1977. *Mathematical Modelling and Digital Simulation for Engineers and Scientists*, Wiley, New York, USA. 332pp.
- Szollosi-Nagy, A., 1976. *Introductory remarks on the state space modelling of water resource systems*. Int. Inst. for Applied Systems Analysis, RM-76-73, 81pp.
- Todini, E., 1996. The ARNO rainfall-runoff model. *J. Hydrol.*, **175**, 339–382.
- Wood, E.F., Lettenmaier, D.P. and Zartarian, V.G., 1992. A land-surface hydrology parameterisation with subgrid variability for general circulation models. *J. Geophys. Res.*, **97**(D3), 2717–2728.

Zhao, R.J. and Liu, X.R., 1995. The Xinanjiang model. In: *Computer models of watershed hydrology*, V.P. Singh (Ed.), Water Resources Publications, Colorado, USA. 215–232.
 Zhao, R.J. and Zhuang, Y., 1963. *Regionalisation of the rainfall-runoff relations* (in Chinese). Proc. East China College of Hydraulic Engineering.

Zhao, R.J., Zhuang, Y., Fang, L.R., Lin, X.R. and Zhang, Q.S., 1980. The Xinanjiang model. In: *Hydrological Forecasting* (Proc. Oxford Symp., April 1980), *IAHS Publ. no. 129*, 351–356.

Appendix A:

Probability-distributed storage models

Analytical solutions of the integrals occurring in the probability-distributed storage model component are presented below for a number of different distribution types.

PARETO DISTRIBUTION

$$F(c) = 1 - \left(\frac{c_{\max} - c}{c_{\max} - c_{\min}} \right)^b \quad c_{\min} \leq c \leq c_{\max}$$

$$f(c) = \frac{b}{c_{\max} - c_{\min}} \left(\frac{c_{\max} - c}{c_{\max} - c_{\min}} \right)^{b-1} \quad c_{\min} \leq c \leq c_{\max}$$

$$\bar{c} = S_{\max} = \int_0^{\infty} (1 - F(c))dc = \frac{bc_{\min} + c_{\max}}{b + 1}$$

$$S(t) = \int_0^{c^*(t)} (1 - F(c))dc = c_{\min} + (\bar{c} - c_{\min}) \left\{ 1 - \left(\frac{c_{\max} - C^*(t)}{c_{\max} - c_{\min}} \right)^{b+1} \right\}$$

$$C^*(t) = c_{\min} + (c_{\max} - c_{\min}) \left\{ 1 - \left(\frac{S_{\max} - S(t)}{\bar{c} - c_{\min}} \right)^{1/(b+1)} \right\}$$

$$V(t + \Delta t) = \pi_i \Delta t - (S(t + \Delta t) - S(t))$$

RECTANGULAR DISTRIBUTION

$$F(c) = \frac{c}{c_{\max} - c_{\min}}$$

$$f(c) = \frac{1}{c_{\max} - c_{\min}}$$

$$\bar{c} = S_{\max} = \int_0^{\infty} (1 - F(c))dc = \frac{c_{\max}((c_{\max}/2) - c_{\min})}{c_{\max} - c_{\min}}$$

$$S(t) = \int_0^{c^*(t)} (1 - F(c))dc = C^*(t) \left[1 - \frac{C^*(t)}{2(c_{\max} - c_{\min})} \right]$$

$$C^*(t) = (c_{\max} - c_{\min}) \left[1 - \left[1 - \frac{2S(t)}{c_{\max} - c_{\min}} \right]^{1/2} \right]$$

$$V(t + \Delta t) = \int_{C^*(t)}^{C^*(t+\Delta t)} F(c) dc$$

$$(i) \quad C^*(t + \Delta t) \leq c_{\max} \quad V(t + \Delta t) = \frac{C^{*2}(t + \Delta t) - C^*(t)}{2(c_{\max} - c_{\min})}$$

$$(ii) \quad C^*(t + \Delta t) > c_{\max} \quad V(t + \Delta t) = \pi_i \Delta t - (S_{\max} - S(t))$$

EXPONENTIAL DISTRIBUTION

$$F(c) = 1 - \exp(-c/\bar{c})$$

$$f(c) = \bar{c}^{-1} \exp(-c/\bar{c})$$

$$S_{\max} = \int_0^{\infty} (1 - F(c)) dc = \bar{c}$$

$$S(t) = \int_0^{C^*(t)} (1 - F(c)) dc = \bar{c} (1 - \exp(-C^*(t)/\bar{c}))$$

$$C^*(t) = -\bar{c} \log(1 - S(t)/\bar{c})$$

$$V(t + \Delta t) = \int_{C^*(t)}^{C^*(t+\Delta t)} F(c) dc \\ = \pi_i \Delta t + \bar{c} (\exp(-C^*(t + \Delta t)/\bar{c}) - \exp(-C^*(t)/\bar{c}))$$

TRIANGULAR DISTRIBUTION

$$F(c) = \begin{cases} 2 \left[\frac{c - c_{\min}}{c_{\max} - c_{\min}} \right]^2 & c_{\min} \leq c \leq \bar{c} \\ 1 - 2 \left[\frac{c_{\max} - c}{c_{\max} - c_{\min}} \right]^2 & \bar{c} \leq c \leq c_{\max} \end{cases}$$

$$f(c) = \begin{cases} \frac{4(c_{\max} - c)}{(c_{\max} - c_{\min})^2} & c_{\min} \leq c \leq \bar{c} \\ \frac{4(c - \bar{c})}{(c_{\max} - c_{\min})^2} & \bar{c} \leq c \leq c_{\max} \\ 0 & \text{otherwise} \end{cases}$$

$$S_{\max} = c_{\min} + \frac{1}{2}(c_{\max} - c_{\min}) = \bar{c}$$

$$S(t) = \int_0^{C^*(t)} (1 - F(c)) dc$$

$$(i) \quad \bar{c} \leq C^*(t) \leq c_{\max} \quad S(t) = \bar{c} - \frac{2(c_{\max} - C^*(t))^3}{3(c_{\max} - c_{\min})^2}$$

$$(ii) \quad c_{\min} \leq C^*(t) \leq \bar{c} \quad S(t) = C^*(t) - \frac{2(C^*(t) - c_{\min})^3}{3(c_{\max} - c_{\min})^2}$$

$C^*(t)$ may be obtained by solving the expression for $S(t)$ above, for example by Newton-Raphson as shown later, or as the solution of a cubic equation in $C^*(t)$.

$$V(t + \Delta t) = \int_{C^*(t)}^{C^*(t+\Delta t)} F(c) dc$$

$$(i) \quad C^*(t + \Delta t) \leq c_{\min} \quad V(t + \Delta t) = 0$$

$$(ii) \quad c_{\min} \leq C^*(t + \Delta t) \leq \bar{c} \quad V(t + \Delta t) = \frac{2(C^*(t + \Delta t) - c_{\min})^3}{3(c_{\max} - c_{\min})^2} - (C^*(t) - S(t))^3$$

$$(iii) \quad \bar{c} \leq C^*(t + \Delta t) \leq c_{\max}$$

$$V(t + \Delta t) = C^*(t + \Delta t) - \bar{c} + \frac{2(c_{\max} - C^*(t + \Delta t))^3}{3(c_{\max} - c_{\min})^2} - (C^*(t) - S(t))$$

$$(iv) \quad C^*(t) + \pi_i \Delta t \leq c_{\max} \quad V(t + \Delta t) = \pi_i \Delta t - (S_{\max} - S(t)).$$

LOGNORMAL DISTRIBUTION

$$F(c) = \Phi \left[\frac{\log c - \zeta}{\sigma} \right] = \frac{1}{2} \operatorname{erfc} \left[\frac{-(\log c - \zeta)}{\sigma \sqrt{2}} \right]$$

where $\Phi(\cdot)$ is the standardised normal distribution function

$$\Phi(x) = \frac{1}{(2\pi)^{1/2}} \int_0^x \exp \left[-\frac{u^2}{2} \right] du,$$

and parameters ζ and σ are the mean and standard deviation of the logarithms of storage capacity.

$$f(c) = \frac{1}{c\sigma(2\pi)^{1/2}} \exp \left\{ -\frac{(\log c - \zeta)^2}{2\sigma^2} \right\}.$$

$$S_{\max} = \int_0^{\infty} (1 - F(c))dc = \bar{c} \exp\left[\zeta + \frac{\sigma^2}{2}\right]$$

$$\begin{aligned} S(t) &= \int_0^{C^*(t)} (1 - F(c))dc \\ &= C^*(t) - \frac{1}{2} \left[C^*(t) \operatorname{erfc}(u_o) - \exp^2\left(\zeta + \frac{\sigma^2}{2}\right) \operatorname{erfc}\left(u_o + \frac{\sigma\sqrt{2}}{2}\right) \right] \end{aligned}$$

$$\text{where } u_o = -\left(\frac{\log C^*(t) - \zeta}{\sigma\sqrt{2}}\right).$$

$C^*(t)$ may be obtained by solving the above expression for $C^*(t)$, for example by Newton-Raphson (as shown later).

$$\begin{aligned} V(t + \Delta t) &= \int_{C^*(t)}^{C^*(t+\Delta t)} F(c)dc \\ &= \frac{1}{2} \left[C^*(t + \Delta t) \operatorname{erfc}(u_1) - \exp\left(\zeta + \frac{\sigma^2}{2}\right) \operatorname{erfc}\left(u_1 + \frac{\sigma\sqrt{2}}{2}\right) \right] (C^*(t) - S(t)) \end{aligned}$$

$$\text{where } u_1 = -\left(\frac{\log C^*(t + \Delta t) - \zeta}{\sigma\sqrt{2}}\right).$$

NEWTON-RAPHSON SOLUTION OF S(t) FOR $C^*(t)$

The Newton-Raphson solution of

$$S(t) = \int_0^{C^*(t)} (1 - F(c))dc$$

for $C^*(t)$ is based on the iterative scheme

$$\theta_{i+1} = \theta_i - f(\theta_i) / f'(\theta_i)$$

where θ_i is the estimate of $C^*(t)$ at the i 'th iteration, and

$$f(\theta_i) = S(t) - \int_0^{\theta_i} (1 - F(c))dc$$

$$f'(\theta_i) = \partial f(\theta_i) / \partial \theta_i = -1 + F(\theta_i).$$

Thus the following iterative solution results

$$\theta_{i+1} = \theta_i - \left\{ S(t) - \int_0^{\theta_i} (1 - F(c))dc \right\} / (F(\theta_i) - 1)$$

where θ_0 is set to the best initial estimate of $C^*(t)$; θ_n is accepted as the best estimate of $C^*(t)$ when the step $(\theta_n - \theta_{n-1})$ becomes sufficiently small.