Journal of Hydrology (2008) 355, 91–105

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

Application of the Kineros2 rainfall–runoff model to an arid catchment in Oman

Aisha Al-Qurashi [a], Neil McIntyre [a,*], Howard Wheater [a], Carl Unkrich [b,1]

[a] Department of Civil and Environmental Engineering, Imperial College London, London, SW72AZ, UK
[b] Agricultural Research Service, United States Department of Agriculture, Tuscon, AZ, USA

Received 22 June 2007; received in revised form 17 January 2008; accepted 7 March 2008

KEYWORDS
Wadi;
Arid;
Rainfall–runoff;
Model;
Oman

Summary The difficulty of predicting rainfall–runoff responses in arid and semi-arid catchments using typically available data sets is well known, hence the need to carefully evaluate the suitability of alternative modelling approaches for a given problem and data set; and to identify causes of uncertainty in order to prioritise research and data. In this paper, we evaluate the distributed model, Kineros2, in application to an arid catchment in Oman, using rainfall–runoff data from 27 storm events. The analysis looks at model sensitivities, uncertainty and performance, based on uniform random sampling of the model parameter space and predictions of features of the observed hydrograph at the catchment outlet. A series of three experiments used different calibration strategies (an 11-parameter calibration, a 5-parameter calibration, and a 3-parameter calibration allowing some spatial variability of the saturated hydraulic conductivity). The parameters most significantly affecting flow peak and volume performance are those controlling infiltration rates on hillslopes. The model output was also generally sensitive to a parameter within the rainfall interpolation model. Relatively little sensitivity to initial catchment wetness was observed. Prediction performance was generally poor, for all events and for all the tested calibration and prediction strategies; and the uncertainty, estimated using model ensembles, was very high. A 2-parameter regression model used in previous work was found to perform better for predicting flow peaks. Literature review shows our results are consistent with experience of other modellers of arid and semi-arid climate hydrology. In order to realise the potential value of distributed, physically based models, for application to arid and semi-arid regions, significant data collection and further research is required, in particular regarding spatial rainfall observation and modelling.
© 2008 Elsevier B.V. All rights reserved.

* Corresponding author. Tel.: +44 207 594 6019; fax: +44 207 594 6123.
E-mail addresses: aisha.al-qurashi@imperial.ac.uk (A. Al-Qurashi), n.mcintyre@imperial.ac.uk (N. McIntyre), h.wheater@imperial.ac.uk (H. Wheater), cunkrich@tucson.ars.ag.gov (C. Unkrich).
[1] Tel.: +1 520 670 6381x178.

0022-1694/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2008.03.022

Introduction

Arid and semi-arid regions pose particular problems for hydrological modellers. Amongst the sources of difficulty are the high temporal and spatial variability of rainfall, the large and variable transmission losses, the seasonal variability of vegetation and its affect on runoff. Furthermore, rainfall and flow observation is often hampered by sparse rainfall and runoff gauge networks (Rodier, 1985; Sorman and Abdulrazzaq, 1993; Al-Qurashi, 1995). These issues add to the role of spatial variability as a factor in rainfall–runoff response, and hence the common view that spatially distributed models are appropriate.

Developments in spatial data sets in the last decade have provided new opportunity to explore the potential applicability of physically based, distributed models to arid and semi-arid regions. For example, digital elevation model and rainfall estimation products are now available with near-global coverage. Furthermore, developments in stochastic modelling techniques allow extensive exploration of model sensitivities and prediction uncertainty; and investigation of alternative calibration procedures and performance measures (Wheater, 2002a). Also, open-source codes for distributed rainfall–runoff models are now available and supported, allowing adjustments to model structures, and coupling with model analysis tools (e.g. the Kineros2 model, Semmens et al., 2008).

However, the fundamental problems of observability of spatially distributed processes and inputs, and lack of model identifiability, severely restrict the practical value of physically based, distributed models (Wheater et al., 1993). Due to generally sparse data, and high spatial variability, applicability to arid and semi-arid catchments is especially questionable (see review below). Nevertheless, predictions of flow are needed in such catchments for flood and water resource management, and use of predictive models is inevitable. Therefore, important questions to ask in the context of arid regions are ''how much performance can we get out of a distributed model, given an available data set''; ''how can we best estimate and illustrate the uncertainty in predictions'', ''when and why do simpler, empirical models perform better?'', ''what can we learn from the model, in order to prioritise data collection?''.

This paper aims to examine the applicability of the physically based, distributed, rainfall–runoff model, Kineros2, to 27 rainfall–runoff events from Wadi Ahin, an arid catchment in Oman. Specific objectives are: (1) To identify the key inputs affecting model outputs, hence indicating important hydrological characteristics and data needs. (2) To test parameter estimation and prediction strategies, in terms of prediction performance. (3) To evaluate and illustrate uncertainty in predictions. (4) To compare results with those achievable using simple empirical analysis. First, we briefly review research into the hydrology of arid and semi-arid regions and implications for modelling.

Hydrology of arid and semi-arid regions

The variability of rainfall is a major consideration when modelling arid and semi-arid zone hydrology. Rainfall characteristics tend to be more variable in space and time compared to humid areas (Pilgrim et al., 1988; Wheater et al., 1991; Al-Qurashi, 1995). Rain gauge densities in Walnut Gulch in Arizona, for example, which are about 1 per 2 km^2, showed highly localised rainfall occurrence with spatial correlations of 0.8 between gauges at 2 km separation, and close to zero at 5–20 km spacing (Wheater, 2002b). Osborn and Renard (1973) recommended 300–500 m separation between gauges to be able to capture localised rainfall. In Saudi Arabia, the typical spacing between rainfall gauges at five experimental basins of a five-year intensive study (Saudi Arabian Dames and Moore, 1988) was 10 km. That study showed that on 51% of rain days, only one or two rain gauges out of 20 recorded rainfall, and the sub-daily data showed even more spotty results. On the other hand, clearly the nature of the rainfall depends on local climate, and may not be so variable. For example, analysis of data from arid areas in Niamy and Niger showed that 80% of total seasonal rainfall was found to fall as widespread events which covered at least 70% of the 100 rain gauge network (Lebel et al., 1997); and studies in semi-arid New South Wales, Australia have also shown spatially extensive, low intensity rainfall (Cordery and Pilgrim, 1970).

Another feature of arid and semi-arid zone hydrology, which has proven difficult to model, is the variability of losses. Various studies have noted the spatial and temporal variability of runoff generation losses. Hughes (1995) discusses the role of non-stationary vegetation cover in controlling infiltration and evaporation losses in southern African catchments, and cites various examples. In situ experiments from the Nahal Zin catchment in Israel reported by Lange et al. (1999) illustrate the spatial variability of runoff infiltration due to soil types, soil crusting and land cover. For example, infiltration rates (after wetting) varied spatially from 5 mm/h on a limestone plateau to 15 mm/h on a sandy crusted plain, to 50 mm/h on a sandy vegetated plain. Casenave and Valentin (1992) report infiltration plot experiments on a wide variety of surface types and conditions in semi-arid West African catchments, finding a similar degree of variability in infiltration capacity. Various studies have indicated large and variable channel transmission losses, and it is not unusual to have high flow in the upper catchment and low or zero flow at the downstream gauge (Renard et al., 1966; Cordery et al., 1983; Walters, 1990; Al-Qurashi, 1995). For example, Sharma and Murthy (1996) analysed 79 events from subcatchments of the Luni River in arid north–west India, finding that transmission losses between upper and lower gauges varied between 8% and 56% of total flow. Hughes and Sami (1992) studied two events in Cape Province, South Africa, and found that 75% and 22% of the flow volume was lost to the alluvium/sand channel bed, respectively. Small-scale studies of channel infiltration include Parrisopoulos and Wheater (1992), who conducted in situ tests in a Saudi Arabian wadi, finding saturated hydraulic conductivities to range between 200 and 600 mm/h within one soil column.

The variability of losses is evident in reported runoff coefficients. Grayson et al. (1992) found the runoff coefficient to vary between 9% and 28% during 3 events in 1997 in the Wagga catchment in Australia. Wheater and Brown (1989) found the coefficient varied from 6% to 80% over 11 events in a Saudi Arabian catchment, and McIntyre et al.

(2007) estimated runoff coefficients to vary from 2% to 38% (excluding two events which had estimated runoff coefficients >100%) from analysis of 36 events from Wadi Ahin in Oman.

The variability of arid zone rainfall and infiltration losses is often been cited as a reason for poor performance of hydrological models. Michaud and Sorooshian (1994a), after modelling runoff in Walnut Gulch (150 km^2), concluded ''Approximately half of the difference between observed and simulated peaks was due to rainfall-sampling errors''. In their case, even data representing a spatial aggregation to the 4 km × 4 km pixel scale generally produced serious underestimations of the peak flows. Wheater and Brown (1989) found wide variation in the unit hydrograph parameters when analysing a catchment in south-west Saudi Arabia (area 597 km^2), which they thought was due to rainfall and transmission loss variability. Grayson et al. (1992) note the fundamental model identifiability problems caused by data limitations, including the variability of rainfall and soil infiltration rates, and found in two catchments that measured infiltration rates needed to be altered by a factor of up to five (to 1000 mm/h in one case) in order to achieve an optimum fit to observed flow. Hughes (1995) notes that representation of transmission losses is a fundamental model limitation, in the context of monthly time-step models in arid and semi-arid catchments.

Other aspects of arid and semi-arid zone climate and hydrology which add to the difficulty of modelling include: the general paucity of data on rainfall, flow, soil properties and initial conditions (Grayson et al., 1992; Nouh, 2006); the influence of seasonal and inter-annual vegetation variability (Hughes, 1995); complexity of channel morphology (Costelloe et al., 2006); loss thresholds associated with over-bank flow (Knighton and Nanson, 1994; Lange, 2005); and for monthly water balance applications, the difficulty of estimating potential evaporation (Hughes, 1995). More extended reviews of arid zone features and implications for modelling are given by Rodier (1985), Pilgrim et al. (1988), Walters (1990), Al-Qurashi (1995) and Wheater (2005).

Description of Kineros and review of applications

Kineros is described comprehensively in Woolhiser et al. (1990) and Smith et al. (1995) and the more recent version, Kineros2, in Semmens et al. (2008). Only an overview of the model is given here. Kineros2 is a physically based, distributed, rainfall–runoff model which is designed for modelling events in arid and semi-arid zone catchments. The catchment is split into a number of rectangular planes (representing hillslopes) and straight-line channels, with a specified connectivity. A set of parameters is specified for each plane and channel. Surface flow is simulated for all planes and channels using a four-point implicit finite difference solution (with an adaptive time-step option) to the kinematic wave equation. Wave movement and depth is controlled by slope, channel geometry, Manning's coefficient (n) and two relief (microtopography) parameters (re and rs). An interception depth parameter (i) specifies an interception loss associated with plant cover. This loss is applied over a specified percentage of the plane area. Soil is represented by either one or two soil layers. Infiltration rate is equal to rainfall rate (following interception loss) until an infiltrability limit is reached. This limit is a function of infiltrated depth, governed by parameters saturated hydraulic conductivity (ks), capillary length scale (g), soil porosity (θ) and a scaling parameter. Surface runoff may be generated by either infiltration or saturation excess mechanisms, although if g is set to zero then infiltration rate is constant at ks. Small-scale variability in ks is included using a coefficient of variation parameter v. There is also a parameter which defines the proportion of soil volume which is rock (r and r_c on hillslope and channel, respectively). A channel base flow may be specified for each channel reach, although this would turn off the channel infiltration in that reach. Other components of Kineros2 represent reservoirs, culverts, urban areas and sediment transport – however, these were not used in our application.

Wheater (1981) used an early version of Kineros to analyse a storm in Wadi Aday, Oman, using a 15-min time-step. The loss was represented by a constant runoff coefficient of 0.75 for areas predominantly covered by hard rock, whereas for the gravel areas a runoff coefficient of 0.50 was used. Manning's coefficients were initially estimated from text-book values but later were calibrated in order to achieve a good fit between observed and modelled peak flow. The calibrated values were 0.025 for the gravel hills, 0.016 for the hard rock areas, 0.032 for a channel reach in a gorge, and 0.030 for the other channel reaches. The author proposed that the results were better than those obtained using the time-area approach because the model is physically based therefore the non-linear routing effects of channel storage and overland flow are represented.

Kineros was applied to Wadi Ahin, Oman by Mott MacDonald (1992) as part of a groundwater recharge analysis. Topographical and geological information was used to determine areas and slopes, and to determine the subcatchment boundaries. The starting values of the ks and θ parameters were determined from field investigation and previous calibrations in the region. Manual calibration was initially done using three storms which were observed at six autographic rain gauges, at a jebel foot flow gauge and at three flow gauges near the coast. At the jebel foot, the calibrated parameter set produced volume errors of +38%, −5% and −44% for the three events, and at the coast the errors were +14%, −13% and +3%. A subsequent calibration on 16 more events, which used assumed 'design' rainfall profiles, had a much wider range of volume errors, but a cumulative error over all 16 events of just 1% at the jebel foot and 6% at the coast. It was considered reasonable to transfer the calibrated Wadi Ahin parameters to adjacent ungauged catchments.

Michaud and Sorooshian (1994b) applied the Kineros model to the 150 km^2 Walnut Gulch catchment, using six rainfall events for calibration and 24 events for validation, using 2-min resolution rainfall data. Initial conditions were estimated using a simpler daily model. The uncalibrated Kineros parameters were estimated from physical observations or from text-book values. n was fixed at 0.04 for all channels and 0.05 for all planes; ks for planes was derived via published relationships with soil classes and varied spatially from 3.61 to 8.92 mm/h, while 42 mm/h was assumed

for channel alluvium. Manual calibration was used, varying *n*, *ks* and the coefficient of variability of *ks* over planes and channels, although the authors were not satisfied that the optimum parameters were found due to the complexity of the calibration task. Performance was considered to be disappointing — for example, the validation root mean square error (RMSE) for peak flows was 79% of the mean observed peak, with four peaks overestimated by more than 100%.

Yatheendradas et al. (2008) applied Kineros2 to eight events from a 6.4 km^2 subcatchment of Walnut Gulch, in order to assess the utility of the model for flash flood forecasting, and to identify the key sources of uncertainty. They used radar rainfall with approximately 1 km × 1 km grid resolution, as well as data from eleven rain gauges, with an input time-step ranging from 4 to 6 min. Twenty-one parameters, two initial conditions and the rainfall were sampled a large number of times (in cases over 100 000) within a Monte Carlo-based sensitivity analysis experiment. They found that the principal cause of uncertainty was the uncertainty in radar rainfall — for example the cumulative rainfall in one event varied between approximately 15 mm and 35 mm depending on the estimation method used. They found that *ks*, *r*, *n*, n_c, and *v* were overall the most important parameters and initial plane wetness *s* was also important in cases. They also demonstrated high uncertainty when trying to predict an event using parameter sets identified from other events.

A range of distributed rainfall–runoff models, each with different process representations, data requirements and methods of incorporating spatial variability, have been applied to arid and semi-arid regions. Examples include El-Hames and Richards (1998), Lange et al. (1999) and Costelloe et al. (2006). These investigators and others report some level of success in simulating the rainfall–runoff response, and demonstrate the potential advantages of distributed modelling. However, it is clear that the fundamental identifiability issues illustrated by Grayson et al. (1992) and Michaud and Sorooshian (1994b), and discussed more generally by Wheater et al. (1993) and Wheater (2002a), persist. Next, we examine the significance of the identifiability of Kineros2, and whether and how it can be managed within the prediction procedure, using an application to Wadi Ahin.

Description of Wadi Ahin

Oman's ephemeral rivers (wadis) may be classified into those that drain to the coast, and those that drain to the interior desert. Wadi Ahin is one of the former, situated in the north of Oman (Fig. 1). The catchment can be subdivided into two areas; Ahin West (upstream of the flow gauging station near Hayl, with elevation above sea level ranging from 300 m to 1300 m, area 734 km^2) and Ahin East (from Hayl to the coast, ranging from sea level to 300 m, area 76 km^2). In this paper we study data from Ahin West (which we refer to as 'the catchment' from here on).

In the piedmont area of the catchment, Hawasinah Nappes, Aruma Group rocks and Tertiary limestones, are interspersed with alluvial wadi deposits. In the upper parts of the catchment the alluvium is typically less than 20 m thick,

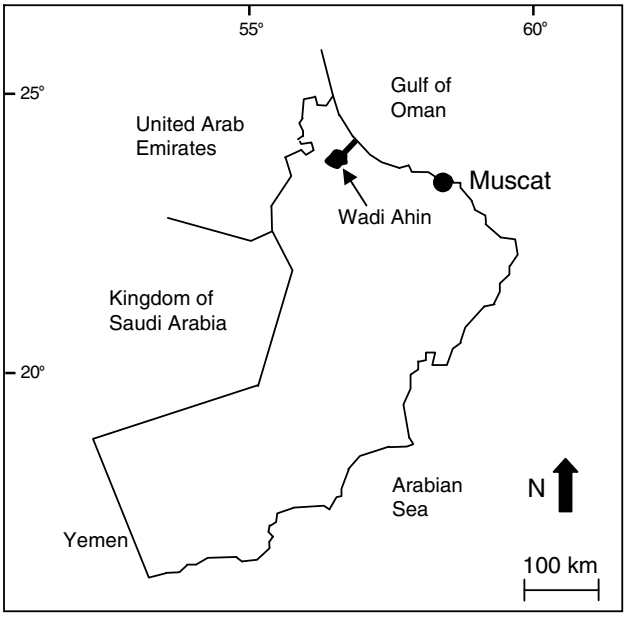

Figure 1 Outline of mainland Oman and the location of Wadi Ahin.

ranges in composition from clay to boulder and in many instances has been weakly to strongly cemented. The wadi channels are mainly gravel and sands. Land cover is mainly desert with sparse vegetation cover. The terrain is mountainous, with slopes ranging from 3% to 92% on hillslopes, and from 0.1% to 0.32% in the channels (measured from a 90 m resolution Digital Elevation Model).

Compared to much of the Arabian Gulf area (see Nouh, 2006), availability of rainfall and flow data for Wadi Ahin is reasonable. There are seven recording rainfall gauges within the catchment (Fig. 2). Five of the gauges operate with digital data loggers. The other two operate using chart recorders. The annual average rainfall at these gauges ranges from 98 mm at Hayl Ashkariyyin (gauge number 7 in Fig. 2) to 143 mm at Al-Waqbah (gauge number 3). The flow gauge at Hayl (location in Fig. 2) is a pressure transducer plus crest stage guages. The stage-discharge curves are based on direct measurement using current meter measurements for low flows, and indirect measurement for peak flows using the slope-area method (a channel survey was done after each flood over a length of 35 m upstream and 32 m downstream of the flow gauge, water surface slopes were estimated from four crest stage gauges in this length, and slope-area method was applied using Manning's equation). In this reach, the channel bed is sand and small cobbles, the right bank is steep rock, and the left bank is a re-cemented conglomerate slope (1:12 gradient on average) with boulders or large cobbles at higher flow depths. A Manning's *n* of 0.036 was used within the slope-area method. For the floods studied in this paper, the top width is estimated to range from less than 5 m to more than 75 m. As the channel bed is shifting, the channel is re-surveyed after each large flood. Events with clearly erroneous flood fluctuations were not included in our data set, although it is noted that human, instrument or stage-discharge errors may affect the quality of the flow estimates.

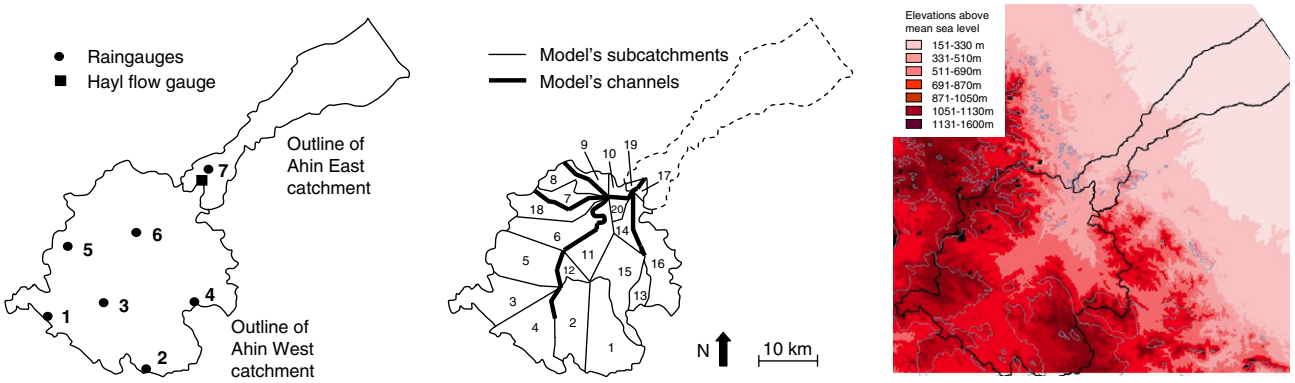

Figure 2 The locations of the seven recording rainfall gauges, schematic of planes and channels for Wadi Ahin application with plane numbers shown, and topography.

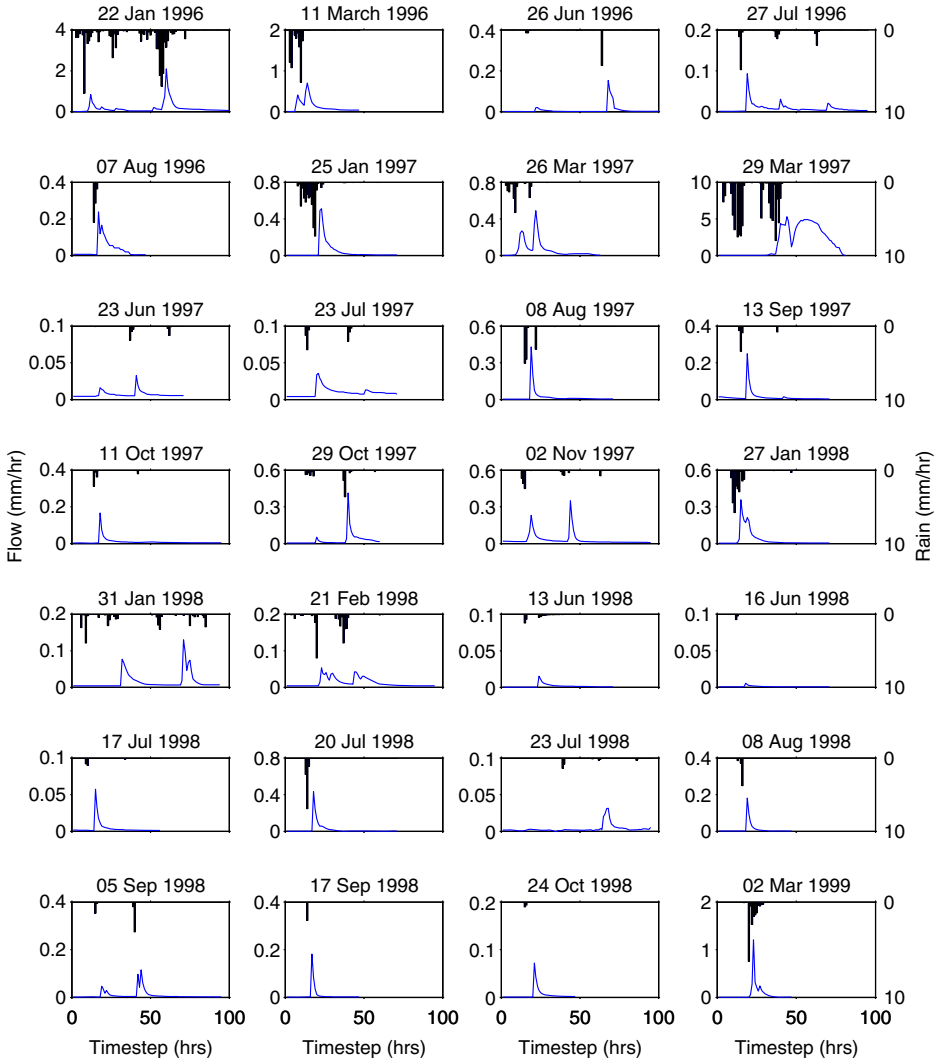

Figure 3 Rainfall and flow time-series for the 27 events (the March 1997 event is split into 26th–28th March and 29th–2nd April, hence there are 28 sub-plots).

Twenty-seven runoff-generating events from 1996 to 1999 are selected, based on availability of the hourly rainfall and flow data, and perceived quality of data. These events are plotted in Fig. 3 (in this plot, rainfall has been averaged without weighting over the seven gauges). Statistics of the events are given in Table 1. Measuring the time lags between peak rainfall and peak flow is complicated because some events contain more than one rainfall and

Table 1 Statistics of the runoff events

Date	Peak rainfall at any gauge (mm/h)	Peak gauge-average rainfall (mm/h)	Total rainfall (sum of gauge-average (mm)	No. gauges with <2 mm rain	Rainfall centroid distance from flow gauge (km)	Base flow (mm/h)[a]	Runoff coefficient	Time lag of peak (h)[b]
22-Jan-96	17	8	61	0	23	0.001	0.29	4, 3 and 3
11-Mar-96	16	6	21	0	20	0.001	0.28	4
26-Jun-96	28	4	5	4	18	0.000	0.13	4
27-Jul-96	16	4	11	0	22	0.001	0.06	4, 3 and 7
07-Aug-96	31	5	9	3	28	0.003	0.15	3
25-Jan-97	12	7	35	0	24	0.000	0.08	4
26-Mar-97	40	8	97	0	25	0.001	1.48	4, 4 and 4[c]
23-Jun-97	8	2	5	3	28	0.004	0.04	4
03-Jul-97	15	3	8	3	27	0.003	0.05	7
08-Aug-97	27	5	13	2	24	0.002	0.09	4
13-Sep-97	17	3	6	2	25	0.003	0.13	4
11-Oct-97	15	2	4	3	19	0.002	0.16	4
29-Oct-97	26	4	9	2	16	0.003	0.15	2 and 2
02-Nov-97	18	3	8	3	16	0.017	0.33	4
27-Jan-98	17	6	22	0	22	0.003	0.09	4
31-Jan-98	22	4	21	0	24	0.003	0.05	9
21-Feb-98	16	6	20	0	24	0.003	0.04	3 and 7
13-Jun-98	4	1	5	4	23	0.000	0.02	9
16-Jun-98	5	1	1	6	19	0.001	0.02	6
17-Jul-98	6	1	2	5	26	0.001	0.09	5
20-Jul-98	22	7	10	3	25	0.001	0.13	4
23-Jul-98	9	1	3	4	20	0.000	0.10	6
08-Aug-98	21	4	5	3	17	0.001	0.09	3
05-Sep-98	8	4	1	6	20	0.001	0.15	4
17-Sep-98	14	2	2	6	19	0.001	0.17	3
24-Oct-98	2	1	1	5	22	0.001	0.26	6
02-Mar-99	23	6	14	4	12	0.000	0.21	3

[a] Baseflow is taken as the flow at the start of the rainfall event.
[b] Where more than one value is given, the values refer to different rainfall bursts/flow peaks within the same event.
[c] These values neglect a major rainfall burst which had no clear associated flow peak (see Fig. 3).

flow peak, and in other cases more than one rain peak can be associated with the single flow peak. However, 37 lags are relatively distinct and lie between 2 and 9 h; 27 of these lie between 2 and 4 h. Eight of the ten largest events in terms of rainfall volume were in the winter months January—April. This period also had relatively uniform rainfall (a greater proportion of the gauges had significant rainfall). There is an exceptional event from 26 March to 1 April 1997. The estimated return period of that peak flow is 85 years, and the runoff coefficient is 1.48, clearly due to under-measurement of rainfall or over-estimation of flow. Excluding that event, the runoff coefficients ranged from 0.02 to 0.33, with a mean of 0.15. The annual average cumulative flow at the Hayl gauge over the four years is approximately 29 Mm^3 (0.92 $m^3 s^{-1}$ or 0.0045 mm/h), falling to 0.9 Mm^3 and 4.4 Mm^3 at two gauges close to the coast. This highlights the degree of transmission loss in parts of the catchment. At the Hayl gauge, the river is dry for 14% of the time, as measured between 1996 and 1999. Base flow is sustained by groundwater springs and, when present, is very small compared to flow peaks (see Table 1). The spatial correlations of hourly rainfall are plotted in Fig. 4.

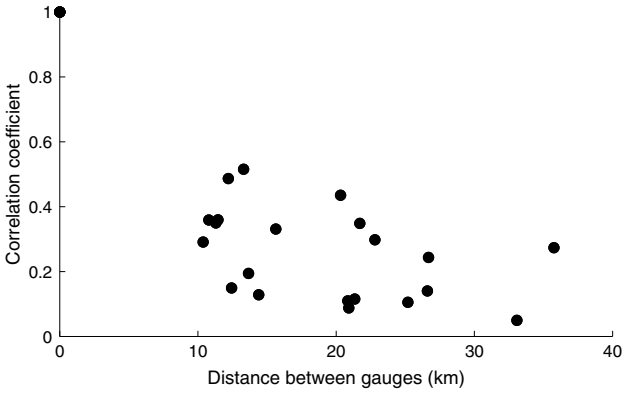

Figure 4 Rainfall correlations between gauges plotted against distance between gauges.

Using substantially the same data set from Wadi Ahin, McIntyre et al. (2007) regressed peak flow, flow volume and runoff coefficient to four rainfall parameters and an antecedent wetness index (base flow), and found that rainfall volume is the primary control on flow volume and peak.

Other effects found to be significant were: increased antecedent wetness increased flow volume and peak; increased spatial variability of rainfall increased flow volume and peak; increased distance of the rainfall centroid from the flow gauge reduced flow volume and peak; increased peak rainfall increased peak flow. In that paper, it was speculated that the performance of the regression model, for predicting flow volumes and peaks at the catchment outlet, might not be improved upon by using a distributed simulation model such as Kineros2 .

Application of Kineros

The application of Kineros2 (which we will refer to simply as Kineros) to the Wadi Ahin data consisted of the following main steps: (1) initial analysis of sensitivity and numerical behaviour; (2) modification of the Kineros code; (3) global sensitivity analysis and parameter estimation (calibration); (4) prediction performance analysis (validation); (5) refinement of calibration procedure, and review of performance improvements. The catchment was split into 20 planes, as illustrated in Fig. 2, with eight associated channel reaches. The planes were based on topography, land cover and soil type. Channels represent the locations of identifiable channels. The slopes and geometries of planes and channels (Table 2) were estimated from the Digital Elevation Model, and the channel cross-section data.

Table 2 Properties of planes and channels

	Length (m)	Width (m)	Area (km^2)	Slope
Planes				
1	15 127	5611	84.4	0.04
2	12 073	5534	66.4	0.03
3	12 459	5163	63.9	0.03
4	10 778	6157	66.0	0.06
5	1789	40 547	72.1	0.35
6	14 060	3884	54.3	0.05
7	11 175	1798	20.0	0.06
8	8520	2882	24.4	0.09
9	695	17 618	12.2	0.92
10	3473	2461	8.5	0.05
11	12 541	3233	40.3	0.04
12	4276	2964	12.6	0.09
13	7851	1907	14.9	0.11
14	19 674	910	17.8	0.01
15	16 305	3933	63.7	0.06
16	18 336	3645	66.4	0.05
17	2616	1298	3.4	0.06
18	12 372	2952	36.3	0.07
19	2710	326	0.9	0.01
20	3429	1630	5.6	0.05
Channels				
1	5526	50	0.3	0.041
2	5887	50	0.3	0.041
3	14 726	50	0.7	0.012
4	12 646	50	0.6	0.055
5	8410	50	0.4	0.073
6	10 302	60	0.6	0.012
7	4618	60	0.3	0.016
8	2922	60	0.2	0.02

Initial analysis

Throughout our analysis, only one soil layer was used. Default parameter values were set to those identified during a previous Kineros analysis of Wadi Ahin (Mott MacDonald, 1992). These parameter values are listed in Table 3. One-at-a-time, local sensitivity analysis around the default parameter values gave a strong initial impression that *ks* and *n* were consistently the most important parameters affecting the simulated hydrograph for every event, and that other parameters and initial conditions had practically zero influence. However, as would be expected, it was noted that the results of this analysis were dependent on the default parameter values about which perturbations were made, especially the value of capillary force, *g* (which had a default value of zero, hence fixing the infiltration rate equal to *ks*).

The local sensitivity analysis indicated some significant volume balance errors. We speculated that volume balance errors may have been due to the restriction, implicit to Kineros, on the maximum number of numerical elements per plane, which might introduce significant errors when representing a long hillslope length by a single plane. This led to a series of experiments on a single plane of length 20 000 m, representing the longest plane used in the Wadi Ahin model. The resolution of the spatial grid was incrementally increased by splitting this single plane into a number of planes in series. A step input of uniformly distributed rainfall was applied with zero infiltration and the Kineros solutions were compared with the analytical solution of Parlange et al. (1981). The Kineros solutions using only one plane adequately matched the analytical solution although minor errors and instability were noted; and adding more planes in series did not always increase accuracy and in some cases significantly increased numerical errors. The volume balance error reported by Kineros was a good indicator of numerical accuracy. Realistic rainfall inputs and *ks* values were used too, also with the conclusion that adding more planes had no significant numerical accuracy benefit.

The outcomes of this initial analysis were: (1) we were confident to proceed with the discretisation in Fig. 2, however volume balance errors need to be monitored; (2) a global sensitivity analysis, using a large number of samples from the parameter space, was needed to generalise the local sensitivity analysis; (3) some adjustments to the Kineros code were needed to allow uniform random sampling of parameters, and to include analysis of rainfall uncertainties and slope.

Modification of the Kineros code

Within the original Kineros code, the rain gauge data are spatially interpolated using the following method. For each model plane (e.g. Fig. 2), a centroid coordinate is specified by the user. At each time-step the ordinates of the rainfall at the three nearest surrounding gauges (or, in special cases, the nearest 2 gauges) are used to define a planar surface, and the average rainfall is computed as the ordinate of

Table 3 Kineros parameters and values

Parameter	Symbol	Units	Default value (planes)	Range (planes)	Default value (channels)	Range (channels)
Manning's coefficient	*n*	$sm^{-1/3}$	0.035	0.01–0.1	0.036	0.01–0.1
Relief (microtopography)	*re*	mm	50	10–100	NA	NA
Relief spacing	*rs*	mm	10	–	NA	NA
Woolhiser coefficient (channel microtopography)	*w*	–	NA	NA	0.15	–
Saturated hydraulic conductivity	*ks*	mm/h	3.7[a]	0–10	41.7	20–50
Capillary length scale	*g*	mm	0	0–500	0	0–500
Variation of *ks*	*v*	–	0.1	–	NA	NA
Initial saturation	*s*	–	0.45	0–0.5	0.45[b]	0–0.5
Soil porosity	θ	–	0.1	–	0.44[c]	–
Interception depth	*i*	mm	2	–	NA	NA
Rock cover	*r*	–	0	–	0	–
Plant cover	*p*	–	0	–	NA	NA
Rainfall interpolation non-linearity	*f*	–	1	0–10	NA	NA
Slope factor	*slp*	–	1	0.5–1	NA	NA

[a] In plane no. 9 the default ks value was 31.22 mm/h, in plane no. 10 it was 20.8 mm/h; in all others it was 3.7 mm/h.
[b] In 4 channel reaches, default *s* value was 0.05; in all other channel it was 0.45.
[c] In channel no. 1, default θ value was 0.15; in all other channel it was 0.44.

this surface over the plane's centroid (Woolhiser et al., 1990). This is formulated into the weighted average of the three gauged values, Eq. (1a), where the three weights sum to one. This weighted average is applied uniformly over the plane. For our Wadi Ahin model, we added a parameter (*f*) into the weighting, Eq. (1b). This parameter controls the weight given to the nearest rain gauge. For example, $f = 0$ means that the three gauges are weighted equally, $f = 1$ is the Kineros default, $f = 10$ weights much more strongly to the nearest gauge. Clearly, this one-parameter model does not aim to produce a realistic spatial rainfall field, but allows a limited exploration of the sensitivity of results to assumptions within the rainfall interpolation.

$$P = \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3 \tag{1a}$$

$$P = \frac{\alpha_1^f P_1 + \alpha_2^f P_2 + \alpha_3^f P_3}{\alpha_1^f + \alpha_2^f + \alpha_3^f} \tag{1b}$$

Additionally, the Kineros code was modified to include a slope factor parameter *slp* which allows the measured slope to be adjusted to an effective slope. Finally, the shell for the code was modified so that Kineros operates within a uniform random sampling procedure.

Global sensitivity analysis and parameter estimation

Global sensitivity analysis, using uniform random sampling of the parameter space, provided some insight into how the Kineros model was behaving, and assisted in deciding which parameters to treat as variables in the calibration process. In order to reduce the sampling problem to a manageable size, θ, *v*, *i*, *w*, *r* and *p* were fixed at the values in Table 3, leaving 11 parameters (for planes: *ks*, *g*, *n*, *re*, *s*, *f*, *slp*; and for channels: ks_c, g_c, n_c, s_c) which were sampled randomly from their uniform distribution defined by the ranges in Table 3. These ranges reflect the range of values in the literature, consistent with the nature of the soils in Wadi Ahin. At this stage, the parameters are not allowed to vary spatially. Initial soil saturation was considered as a parameter due to absence of measurements. Its range was limited to 0–0.5 because values greater than 0.5 often caused numerical instability problems. Baseflow was set to zero in all channels except the outlet channel (channel 8 in Table 2) for which the baseflow values in Table 1 were applied. This treatment of baseflow allowed Kineros to simulate channel infiltration for over 95% of the channel length, and also to include baseflow in the catchment outlet flow estimate.

A total of 20 000 parameter sets were sampled and the model was run using each set, for each of the 27 events. Five outputs from Kineros were recorded for each run: peak flow (Q_p), flow volume (Q_v), time to peak after beginning of simulation (Q_t), total rainfall volume (R_v), numerical mass balance error (*E*). Five performance measures (objective functions) were then calculated, as specified in Eqs. (2)–(6). Q'_p, Q'_v and Q'_t are the observed values at the Hayl flow gauge. In OF_3, K_t is fixed to a representative time-scale of 4 h (the most common time lag between peak rainfall and peak runoff – see Table 1).

$$OF_1 = \left|\frac{Q'_p - Q_p}{Q_p}\right| \tag{2}$$

$$OF_2 = \left|\frac{Q'_v - Q_v}{Q_v}\right| \tag{3}$$

$$OF_3 = \left|\frac{Q'_t - Q_t}{K_t}\right| \tag{4}$$

$$OF_4 = \frac{1}{2}(OF_1 + OF_2) \tag{5}$$

$$OF_5 = \frac{1}{3}(OF_1 + OF_2 + OF_3) \tag{6}$$

In five events flow gauging ceased before the end of the runoff recession so Q_v does not necessarily represent the full runoff volume, however Q'_v is always measured over

the same time-period as Q_v. A sixth objective function measures the average performance in terms of flow peak and volume over a specified N events

$$OF_6 = \frac{1}{N}\sum_{i=1,N} OF_4 \quad (7)$$

The response of the objective functions over the range of each parameter defines the sensitivity of that objective function to that parameter. In particular, we are interested in the response over the parameter sets which give a 'good' result (the others are considered to be outside the relevant parameter space). Due to significant scope for observed flow errors, a 'good' result is defined here as an objective function value less than 0.3. As a measure of the relative univariate sensitivities, we use the univariate Kolmogorov–Smirnov (KS) statistic. For each parameter, this measures how far the distribution of the good parameter values deviates from the uniform distributions defined in Table 3. The method of calculation of the statistic is described in McIntyre et al. (2005).

The KS statistic was calculated for each parameter, for each event, for OF_1, OF_2 and OF_3. The parameters were ranked in terms of the significance of the KS value for each event. The average rank over all 27 events, for each objective function, is presented in Table 4, together with an overall rank based on the average of these averages. In the case of OF_1, OF_2 and OF_3, respectively, two, two and six events were omitted in the calculation of average ranks because there were few (less than 100) parameter sets giving good performance. *ks* is found to be the most influential parameter overall, followed by *g*, *n*, n_c and *f*, and there were only small variations in rank between objective functions. All parameters were significant at the 95% level. Conclusions about sensitivity were unchanged when using an alternative sensitivity measure (based on comparing variance of the good parameter values with the prior variances) and when visually reviewing parameter-objective function scatter plots. Fig. 6 shows the scatter plots of the 11 parameters against the values of OF_6, indicating that, except for *g* and *ks*, the optimal values are not well-identified by the 20 000 samples. Fig. 6 also illustrates that the response of OF_6 seems to suggest that *re* is more important than the summary statistics in Table 4 suggest; and that the event-averaged performance was poor — at best 0.52 (52% error).

Table 4 Average ranking of importance of parameters to OFs, and overall rank

	OF_1 (Q_p)	OF_2 (Q_v)	OF_3 (Q_t)	Overall rank
n	3.5	3.6	3.3	3
re	6.9	6.0	6.4	6
ks	1.2	1.1	1.9	1
g	3.0	2.6	3.8	2
s	6.3	6.7	6.5	7
n_c	4.8	6.0	2.3	4
ks_c	8.2	8.4	9.0	9
s_c	8.8	9.3	9.5	10
g_c	9.4	9.0	9.9	11
f	6.3	5.4	5.1	5
slp	7.5	7.8	8.3	8

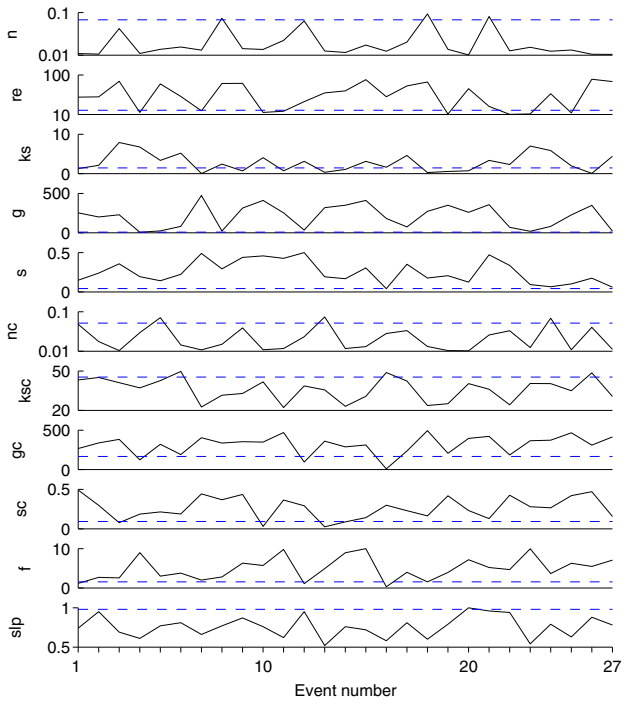

Figure 5 Optimum parameter values over events (OF_4). Dashed line shows the parameter set which was optimal over all events combined. Parameter units are given in Table 3.

Although the rainfall parameter *f* was ranked only 5th in terms of its effects on flow, its importance to rainfall volume was large; variation of *f* from 0 to 10 caused between 1% (22 January 1996) and 156% (17 July 1998) variation in rainfall volume, and on average 31%.

The best out of all the samples, as measured by the minimum value of a chosen OF, gives an approximation to an optimum parameter set. The optimum parameter values were found, in general, to vary widely over the events and over the OFs. For example, Fig. 5 shows the variation of the OF_4 optimal parameter values over the 27 events, and (in dashed lines) the single parameter set which produced the minimum value of OF_4 averaged over all events (OF_6). Only *ks* may be considered as reasonably consistent over events. No significant correlations between event type (spatial variability, magnitude, location or intensity of rainfall) and parameter values were observed.

The parameter interaction is indicated by correlation coefficients. The correlations between parameters were calculated for each event, based on the 'good' parameter sets (omitting the events where these do not exist) as defined by $OF_4 < 0.3$. Strong negative correlations consistently exist between *ks* and *n*, *ks* and n_c and *ks* and *g*; strong positive correlation consistently exist between *ks* and *re*, and *ks* and *f*. The magnitudes of these correlation coefficients were consistently greater than 0.1 over the 27 events; on average the values were −0.34, −0.55, −0.15, 0.23 and 0.12, and at greatest −0.58, −0.80, −0.42, 0.41 and 0.37, respectively. For each event there were various other strong correlations, but these were not consistently strong over events. For OF_1 and OF_2 (using 0.3 as the threshold) the results were comparable. For OF_3, *n* and n_c were also

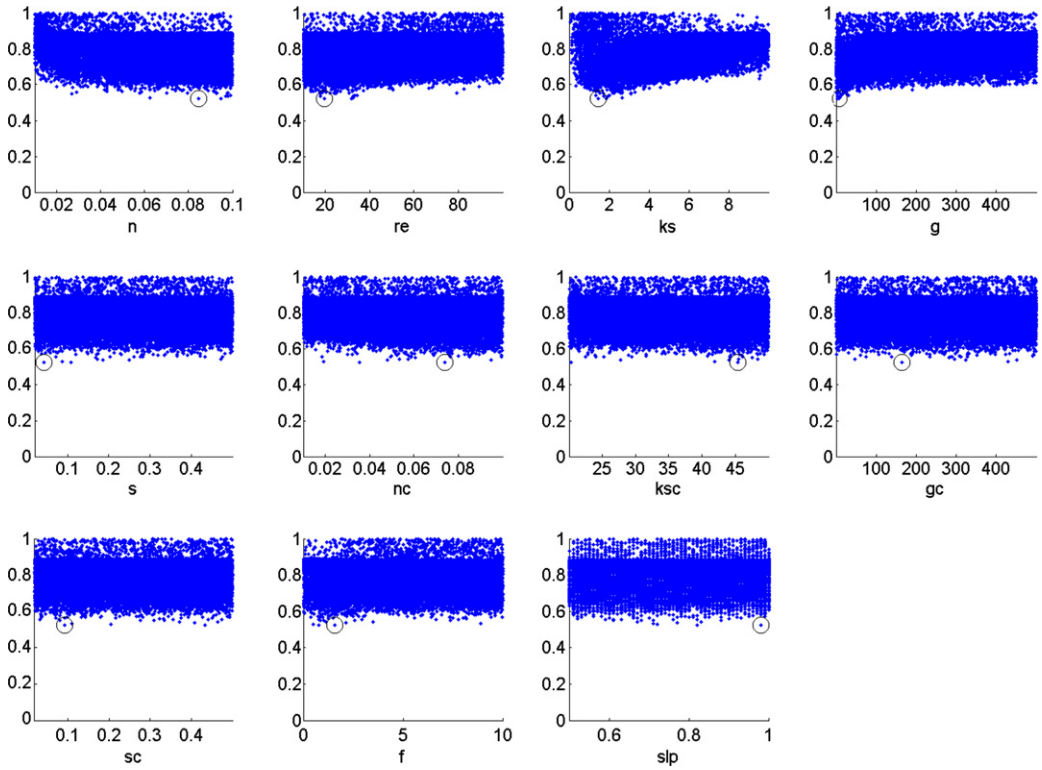

Figure 6 Scatter plot of parameter values against OF_6. Parameter units are given in Table 3.

significantly correlated (with an average correlation coefficient over all events of −0.20 and minimum −0.85).

The numerical volume balance error reported by Kineros varied between 0% and 59% over all events, although only four events resulted in any errors greater than 20%, the overall average was 0.50%, and the 'good' parameter sets ($OF_4 < 0.3$) gave an average error of 1.38%, and the best parameter sets (minimum OF_4) gave an average error of 1.62%. Interestingly, this implies that the realistic model responses are, on average, more numerically problematic than the rest of the responses. The number of failed runs (i.e. when Kineros crashed) varied between 0 and 18 of the 20000 runs, and on average was 6.

The outcomes from this global sensitivity analysis are: (1) The primary importance of the Manning's coefficient, n, the capillary length, g, and the ks are confirmed; the channel roughness, n_c, and the rainfall parameter, f, are the next two important overall. (2) The variability of the estimated optimum parameter values is high over events and this appears to be random, except for optimum values of ks which were relatively consistent over events. (3) There are high correlations between some of the important parameters, with ks and g, and n and ks having especially strong interaction, and this partly explains the variability of optimum parameters over events. (4) Numerical errors are high for some parameter sets, but less than 2% on average for the best parameter sets.

Prediction (validation) performance analysis

The best fit from the 20 000 sampled parameter sets is recorded using objective functions OF_1 to OF_5, Eqs. (2)–(6) for each event; and the best fit averaged over all 27 events is recorded using OF_6, Eq. (7). This gives an approximation to the best achievable performances using Kineros. Then, the following four prediction strategies were tested:

(1) As a benchmark, the default parameter set (based on Mott Macdonald, 1992) is applied to each event and the performance is recorded.
(2) Each event in turn is disregarded in the computation of OF_6, and the subsequent optimum parameter is applied to that event. Thus, each of the 27 events is predicted using the parameter set which produces the overall best fit to the other 26 events (i.e. validation).
(3) The optimum parameters sets derived for each of the other 26 events are used individually to produce an ensemble of predictions for each event. Taking the average of the ensemble is the third prediction strategy.
(4) Based on the same ensemble, a second average is recorded, where only the parameter sets which achieved objective function value less than 0.3 during calibration are allowed to contribute to the ensemble, rather than all 26.

Prediction performances achieved using these four approaches (default parameter set, calibration by lumping all events into one objective function, ensemble model average and reduced average) are compared. Our analysis of prediction performances concentrates on the use of OF_4 and OF_6, as combined measures of flood and volume performance, however other measures are also reported for comparison.

The minimum values of OF_1 achieved in calibration were always less than 0.01 (i.e. at least one of the 20000 parameter sets was able to fit the observed flow peak to within 1%) except for one event which achieved a match to within 3%. The minimum values of OF_2 were all practically zero, except for the extreme flow event of 26 March 1997 where, at best, volume was under-predicted by 40% (the observed runoff coefficient was 1.48). The minimum value of OF_3 was practically zero for 21 events, and close to zero (less than 0.12) for the other 6.

The best values of OF_4 from calibration, and those achieved using each of the first three prediction (validation) strategies listed above, are shown in Fig. 7. This shows that using the parameter set identified by lumping the performance over 26 events into one objective function (OF_6) is the safest with an average validation performance $OF_4 = 0.60$, while the default parameter set gave more 'good' results with $OF_4 < 0.3$ (4 good results). Using the third and fourth strategies produced 5 and 6 good results, respectively, but performances were much more variable. The variable performance of the ensemble average may be due to the smoothing of peak flows, as timing of peaks varies within the ensemble. As an example illustration of the fits of the various calibration objective functions and prediction strategies, Fig. 9 shows time-series results for the event of 22 January 1996. Fig. 9 also indicates that the lower objective function values do not always produce the most visually pleasing fits.

Using OF_1 as the calibration and the prediction performance criteria did not lead to markedly different conclusions. The single best parameter set identified by averaging OF_1 over the events produced an average performance in validation of 0.78 and 6 good results ($OF_1 < 0.3$), while the default parameter set had corresponding values of 0.60 and 5, and the ensemble average achieved corresponding values of 1.07 and 7. Using the OF_2-optimal parameter sets, the corresponding flow volume performances were 0.45 and 9; 0.73 and 9; and 1.11 and 11.

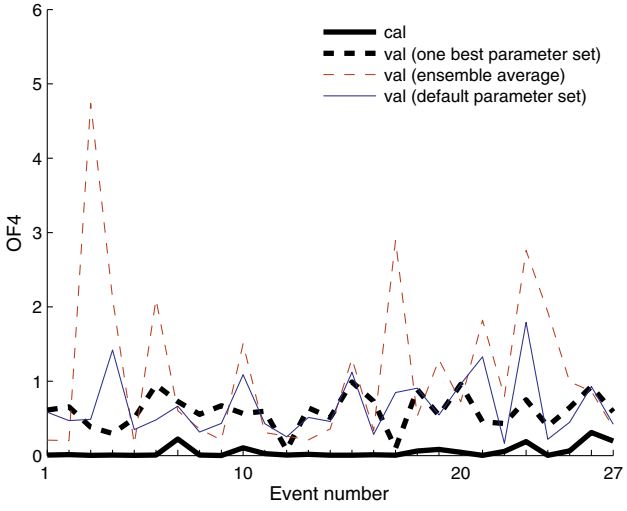

Figure 7 Performance (OF_4) in calibration and validation using different prediction strategies, for each of the 27 events (11-parameter calibration).

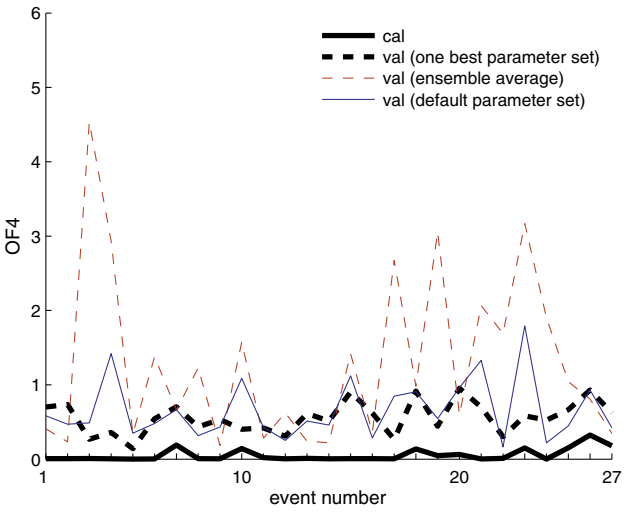

Figure 8 Performance (OF_4) in calibration and validation using different prediction strategies, for each of the 27 events (5-parameter calibration).

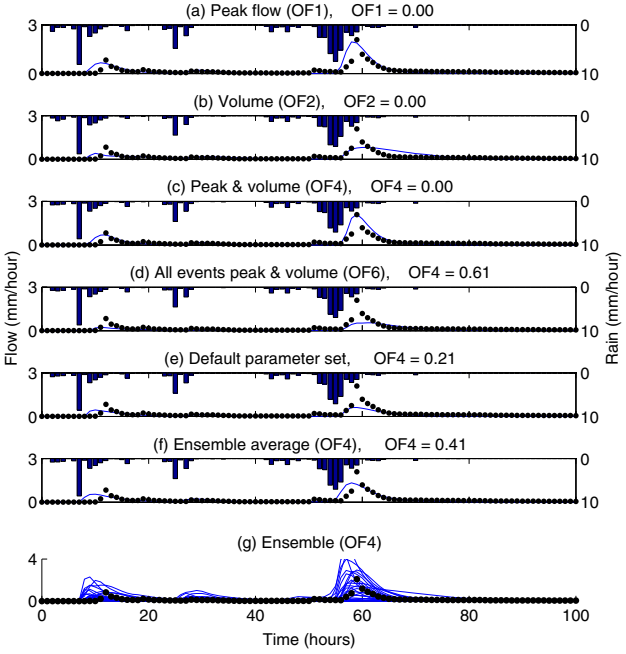

Figure 9 Example of calibration and validation results for 22 January 1996. Bars from top axes are rainfall; dots are observed flow data, and lines are simulated flow.

The ensembles fully encompassed the observed data for all events with one exception (the March 1997 event), and therefore robustly represent the prediction uncertainty. However, the uncertainty is unacceptably large for practical purposes. For example, the ensemble of predicted runoff coefficients, using the OF_4 optimal parameter sets, varied on average from 0.04 to 0.79, and in one case (25 January 1997) from 0.01 to 0.91. This degree of variability is evident in the example in Fig. 9g. The runoff coefficients simulated by any one of the 27 parameter sets varied widely over the events, on average from 0.03 to 0.54, and in one case (using

the parameter set which was optimal for 16 June 1998) from 0 to 0.72. This confirms the strong dependence of the model response on the spatial–temporal nature of the rainfall, as well as on the selected parameter set.

In these predictions, which transfer parameter sets between events, or use a single parameter set identified over 26 events, the initial saturation parameters (s and s_c) and the rainfall distribution parameter (f) are implicitly assumed to be transferable from one event to another. This is debatable. In principle they are clearly not transferable, however, as the values of the other parameters are dependent on the values of s, s_c and f, arguably the sampled optimal values of these three parameters should not be viewed independently from the associated parameter set. In order to test whether this makes a difference in practice, we repeated the prediction exercise while fixing s, s_c and f at the values identified for the event being predicted rather than those inherent to the parameter set under trial. This did not notably change prediction performance: for example, the average OF_4 using the OF_6 calibration increased from 0.60 to 0.64; and 3 and 7 good results ($OF_4 < 0.3$) were achieved using the default parameter set and the ensemble average, respectively.

Refinement of calibration procedure, and review of performance

The procedure for estimating the optimal model parameters, described above, was rather ambitious, with respect to the number of parameters to be optimised (11). In particular, it seems likely that the parameter variability and poor prediction performances may have been associated with the over-parameterisation, and the associated sparseness of the uniform random sampling of parameters. The problem was reduced by eliminating channel infiltration, and fixing $g = 0$ so that hillslope infiltration rate is constant at ks; and fixing $slp = 1$ hence relying on the variation of the re microtopography parameter to optimise slope scaling effects. Calibrating the five remaining active parameters (n, n_c, ks, re and f), the model evaluation was repeated using the same procedure. Results for OF_4 are in Fig. 8. Using OF_4, the performance using the single best (OF_6) parameter set was 0.56 on average, with 4 good performances; the equivalent values using the ensemble method were 1.33 and 5. Using the OF_1-optimal parameter sets, the corresponding peak flow performances were 0.60 and 5; and 1.29 and 6. Using the OF_2-optimal parameter sets, the corresponding flow volume prediction performances were 0.43 and 12; and 1.28 and 9. In summary, the 5-parameter version of Kineros did not significantly change prediction performance (using best values, 6 good predictions compared to 7 for OF_1; 12 good predictions compared to 11 for OF_2; 5 good predictions compared to 6 for OF_4). The variability of the optimal value of ks and n_c over events did not reduce from the results shown in Fig. 6, hence there was no apparent advantage in terms of parameter identifiability.

Finally, spatial variability was introduced into the most sensitive parameter, ks. To represent the different soil types, the ks values for plane numbers 13, 14 and 15 were allowed to vary independently from the value used for the other planes. This was based on the soil types and the geology. The other most generally important parameters, n and n_c, were also calibrated, but were set to be uniform over the catchment. The other parameters were fixed at their optimum value from the previous calibration. Only 25 events were run in this case (the 22 January 1996 and 26 March 1996 events failed to run due to unknown numerical problems). The performances over the 25 events did not generally improve upon the 5-parameter version. For example, using the parameter set which was OF_6-optimal, gave an average prediction OF_4 value of 0.61 with 3 good results.

Discussion

A global sensitivity analysis of 11 parameters found that they were all significantly affecting all objective functions over all tested events. This was foreseen, as we knew that they are all influential in theory. The finding that the parameters governing hillslope infiltration rates, ks and g, and the Manning's roughness, n, are consistently the most important confirms previous research in arid and semi-arid catchments (e.g. Sorman and Abdulrazzaq, 1993; El-Hames and Richards, 1998; Yatheendradas et al., 2008). Although channel infiltration is thought to be significant in practice, the parameters governing channel infiltration were found to be the least significant, as measured by the KS statistic. Perhaps this was because of the limited channel bed area compared to the planes in the model (Table 2); an extended channel network would be expected to increase importance of channel bed infiltration (e.g. Grayson et al., 1992). Initial saturation of channel soil was the second least significant parameter, and that of the planes was only seventh out of 11 parameters. However the initial saturation was varied only over the range 0–0.5 because of numerical stability problems at higher values, although we think this range is not unreasonable for the Wadi Ahin events. Correlation between some parameters was strong, particularly between ks and the other relatively important parameters, for example the roughness parameter n. This, we assume, is because the main function of n is to control flow velocity and residence time, and hence cumulative infiltration loss. Correlations which we expected (e.g. slope parameter slp and roughness n have opposite effects in theory) were less strong, although were still significant.

The difficulty of identifying a model which produces acceptable results over different events has been demonstrated. A key parameter, ks, was found to be relatively constant over events (Fig. 5), however other key parameters, notably g and n, are much more variable. Furthermore, transferring a parameter set from one event to another usually produced very poor performance. We were not able to find evidence that the event-optimal parameter values were related to the spatial location and variance of the rainfall. There was limited evidence that the more spatially varied the rainfall, the worse the calibration performance (e.g. the correlation between OF_4 and the spatial variability index in Table 1 = 0.43), but no evidence of the same for prediction performance. We also tried eliminating the parameter sets derived from events with highly variable rainfall, on the speculation that these are especially difficult to use for calibration, however this was not beneficial.

Michaud and Sorooshian (1994b) did a comparable application of Kineros to Walnut Gulch, but with considerably better coverage of rain gauges and with higher time-resolution data, better spatial data on soils, higher resolution elevation models, and the catchment area was approximately one fifth of that of Wadi Ahin. They found that the validation root mean square error (RMSE) for peak flows was 79% of the mean observed peak, with four out of 24 predicted peaks overestimated by more than 100%. Our best validation peak flow RMSE error (achieved using our 5-parameter calibration to OF_1) was 147% of the mean observed peak, and none of our 27 peaks were overestimated by more than 100%, although at best only 7 of our peaks were predicted to within 30%. Their validation RMSE for volumes was 73% of mean observed volume, while our equivalent was 299% (achieved using our 5-parameter calibration to OF_2), although this reduces to 106% if the March 1997 event is omitted. Given the much poorer spatial data available for Wadi Ahin, the comparison with the results from Walnut Gulch (Michaud and Sorooshian, 1994b) is not disappointing, and might be attributed to the less localised rainfall in Wadi Ahin (compare the correlations in Fig. 4 with the Walnut Gulch correlations mentioned in our literature review).

We previously applied regression models to predict flood peaks and volumes in Wadi Ahin (McIntyre et al., 2007), using the substantially the same events as presented in the current paper (the same events but splitting some of the longer events into two events, incorporating one event which would not run in Kineros due to numerical problems, and excluding the outlying March 1997 event to give a total of 34 events in the regression). We found that, by linear regression of flow peak and volume against gauge-average rainfall, 16 of 34 observed flow peaks were predicted to within 30%, and 11 of 34 observed flow volumes were predicted to within 30%. Using Kineros, our best corresponding results were seven out of 27 and 12 out of 27. A potential benefit of a distributed simulation model over a lumped empirical model is the theoretical ability to significantly improve predictions by accounting for the spatial–temporal nature of the rainfall and rainfall–runoff processes, however in our application Kineros has failed to do so.

The reasons for the limited performance of Kineros in the Wadi Ahin application are not proven. A starting point for further analysis would be more extensive representation of rainfall errors in the experiments. The one-parameter rainfall model provided a limited exploration of the possible realisations of spatial rainfall. The use of a space-time stochastic rainfall model for generation of rainfall scenarios would allow more extensive analysis of sensitivities of Kineros, and could be used to demonstrate that rainfall estimation problems are over-riding, which cannot be concluded from this paper, although is indicated by Yatheendradas et al. (2008). There is likely to be significant bias in the estimate of areal rainfall because the gauges in Wadi Ahin are biased towards lower elevations, and this may be the cause of the estimated runoff coefficient of 1.48 for the March 1997 event. A stochastic model which includes elevation adjustment is therefore recommended. Chandler and Wheater (2002) and Yang et al. (2005) provide such a model for daily rainfall; for sub-daily rainfall further development of continuous space-time Poisson process models is required (Wheater et al., 2005).

The flow data may also be questioned, most notably that of the March 1997 event. Fig. 3 shows that this flow response has an uncharacteristic high flow duration, and there is no clear flow peak associated with one of the major rainfall bursts. The quality of this data was not investigated as part of this study, but is recommended. The sensitivity of validation results to omitting this event from the calculation of OF_6 was tested, and found not to improve performance as measured by Eqs. (2)–(6). Infact, in terms of the number of events for which $OF_4 < 0.3$ was achieved, it marginally degraded performance.

Using an ensemble of 26 parameter sets estimated from different events has been useful as it allows us to visualise the degree of uncertainty in predictions (e.g. Fig. 9g), and potentially it provides a basis from which to investigate how much the uncertainty may be reduced, for example by improved rainfall estimation. It also provides a framework for adding sources of uncertainty, for example the uncertainty which may arise from introducing a more comprehensive set of plausible rainfall inputs or more than one-parameter set from each event. At present the uncertainty is unsatisfactory – so high (with runoff coefficients within each ensemble prediction ranging on average from 0.04 to 0.79) that it is practically useless for decision-support. This conclusion is consistent with the simultaneous finding of Yatheendradas et al. (2008) in their application of Kineros to flash flood forecasting in Walnut Gulch.

Our Kineros application, including several iterations of the experiments reported in this paper, has cost more than 18 person-months, and more than the equivalent of 12 weeks of continuous processing on a desktop computer. In contrast, our regression analysis of the same data (McIntyre et al., 2007) took no more than six person-days, and no more than 1 h of processing time. This is an unfair comparison, because of the research context and because simply using the Mott Macdonald (1992) Kineros parameters would have significantly saved us time without much loss of prediction performance (see Figs. 7 and 8). However the differences in complexity of the modelling approaches and corresponding resource cost are clearly wide, yet the differences in prediction performances and the insights into key processes seem to be small, or in favour of the simpler approach.

Kineros has the underlying restriction that the spatial variability of slope, and soil and surface properties are lumped to plane scale/channel scale effective values. Therefore, especially where plane areas and channel lengths are high (as in our Wadi Ahin application), the model may not be considered truly distributed. Hence, the ability to model rainfall–runoff processes in a distributed, physically based manner has not been tested in this paper. Whether this would be theoretically possible using a different model and a much finer spatial discretisation is arguable, due to fundamental commensurability and observability issues (Beven, 1989; Wheater, 2002a). Although the available digital elevation data would support a much finer resolution model of the surface processes, the uncertainty arising from the other, less observable surface properties, subsurface properties and spatial rainfall properties seems likely to remain limiting. In the few cases where research-quality data sets are available, and much more spatially detailed models have been used, calibration

of infiltration and roughness parameters remains advisable and performance remains quite poor (e.g. Grayson et al., 1992; Michaud and Sorooshian, 1994b). An additional constraint on the physical realism of our application of Kineros is that the channel network is limited to the major channels (Fig. 2), and the omission of the more distributed stream network is expected to cause significant underestimation of total channel transmission loss (e.g. Grayson et al., 1992). Instead, within the calibration process, these losses were lumped into plane infiltration. Finally, as with all models, Kineros is based on simplified representations of processes, for example neglecting lateral subsurface flow, soil crusting effects and flow dependence of n, and may benefit from structural development for any particular application, should supporting observations be available.

The value of Kineros is questionable in the context of this Wadi Ahin application, however it could have three advantages in different circumstances. Firstly, it contains parameters which, in design at least, are physically based and in some cases may be related to soil classifications (Michaud and Sorooshian, 1994b). Therefore it may be applicable to represent physical change, or for catchments with relatively good spatial data on soil properties (although this was implied to be of limited value by Michaud and Sorooshian, 2004b; also see Grayson et al., 1992). Secondly, the ability of Kineros to represent distributed runoff processes and generate spatially continuous runoff may be attractive for some applications, for example simulation of the effects of local interventions such as storage reservoirs or inundation mapping, whereas we only aimed here to predict flow variables at the catchment outlet. Thirdly, the model generates temporally continuous runoff within each event, while we only assessed prediction performance for volumes and peaks of the hydrographs (although the full hydrograph can sometimes be predicted from the peak and/or volume (e.g. Sharma and Murthy, 1996), and this is somewhat evident in Fig. 3).

Conclusions

The distributed, physically based rainfall–runoff model, Kineros was applied to 27 rainfall–runoff events (Fig. 3) from an arid catchment in Oman, with the main objective of assessing the value of the model for predicting features of the event hydrographs at the catchment outlet. A random sampling experiment was applied to evaluate the global sensitivity of parameters and to approximate the optimal parameter sets, using objective functions representing peak flow, flow volume and time to peak performances. The factors found to most affect the volume and peak performance were generally consistent with those previously identified in the literature – infiltration rates in the hillslopes, the Manning's roughness in hillslopes and channels, and the rainfall parameter. When each event was treated independently, using the best of the sampled parameter sets, Kineros was able to accurately simulate flow peak, flow volume and time to peak, for almost all events. However, the parameter sets which were estimated to be optimal for individual events did not perform well when transferred to other events. Furthermore, the parameter set which gave best calibration performance over any combination of 26 events did not generally produce acceptable performance (defined as within 30% of observed) when used to predict the 27th event (Figs. 7 and 8). The parameter sets identified for the events individually were used to produce an ensemble of predictions. This produced very high uncertainty, for example runoff coefficients within each ensemble ranged on average from 0.04 to 0.79. The average of the ensemble was a very poor representation of the observed flow data. Different objective functions (peak and volume, peak only, and volume only) were tried, as well as a second experiment where the infiltration model was substantially simplified, and a third experiment which allowed spatial variability in the infiltration rate, but the same general conclusions were reached. It was concluded that the two-parameter regression model used by McIntyre et al. (2007) was more successful than Kineros at predicting flow peaks and was slightly worse for predicting flow volumes at the catchment outlet.

The potential value of distributed (or semi-distributed) rainfall–runoff models in application to arid regions lies mainly in their ability to represent surface topography and/or rainfall in a spatially distributed manner; and to simulate spatially distributed runoff. The Wadi Ahin application reported in this paper supports the opinion that data sets typically used for distributed (or semi-distributed) rainfall–runoff modelling in arid regions cannot provide an accuracy which justifies the effort and expense of this modelling approach. The limitations imposed by relatively sparse observations of rainfall are of particular concern, as well as the need for calibration of key surface and subsurface parameters. Further performance enhancements might be achieved for Wadi Ahin by using a more spatially distributed application of the Kineros model, which could fully exploit available topographical data. However, the value of making the model even more complex in light of the general data restrictions is questionable and is not recommended. Major research priorities, towards improved hydrological models for arid regions, are improved methods for rainfall observation and modelling.

Acknowledgements

The authors would like to thank the Ministry of Regional Municipalities, Environment and Water Resources of the Sultanate of Oman for provision of the Wadi Ahin data; the Islamic Development Bank, who have partly funded this research; and two reviewers for useful remarks.

References

Al-Qurashi, A.M., 1995. Rainfall–runoff relationships in arid areas, MSc thesis, Heriot-Watt University, Edinburgh, UK.

Beven, K., 1989. Changing ideas in hydrology-the case of physically-based models. Journal of Hydrology 105, 157–172.

Casenave, A., Valentin, C., 1992. A runoff capability classification system based on surface features criteria in semi-arid areas of West Africa. Journal of Hydrology 130, 231–249.

Chandler, R.E., Wheater, H.S., 2002. Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. Water Resources Research 38 (10), 1192. doi:10.1029/2001WR00090.

Cordery, I., Pilgrim, D.H., 1970. Design hydrograph methods of flood estimation for small rural catchments. Civil Engineering Transaction, CE12, 2 (Australia).

Cordery, I., Pilgrim, D.H., Doran, D.G., 1983. Some hydrological characteristics of arid western New South Wales. In: Hydrology

and Water Resources Symposium 1983, Inst. Engrs. Austral., National Conference Publ. No. 83/13, pp. 287—292.

Costelloe, J.F., Grayson, R.B., McMahon, T.A., 2006. Modelling streamflow in a large anastomosing river of the arid zone, Diamantina River, Australia. Journal of Hydrology 323 (1—4), 138—153.

El-Hames, A.S., Richards, K.S., 1998. An integrated, physically based model for arid region flash flood prediction capable of simulating dynamic transmission loss. Hydrological Processes 12, 1219—1232.

Grayson, R.B., Moore, I.D., McMahon, T.A., 1992. Physically based hydrologic modeling 1. A terrain-based model for investigative purposes. Water Resources Research 28 (10), 2639—2658.

Hughes, D.A., Sami, K., 1992. Transmission losses to alluvium and associated moisture dynamics in a semi-arid ephemeral channel system in southern Africa. Hydrological Processes 6, 45—53.

Hughes, DA., 1995. Monthly rainfall—runoff models applied to arid and semiarid catchments for water resource estimation purposes. Hydrological Sciences Journal 40 (6), 751—769.

Knighton, A.D., Nanson, G.C., 1994. Flow transmission along an arid zone anastomosing river, Copper Creek, Australia. Hydrological Processes 8, 137—154.

Lange, J., 2005. Dynamics of transmission losses in a large arid stream channel. Journal of Hydrology 306 (1—4), 112—126.

Lange, J., Leibundgut, C., Greenbaum, N., Schick, A.P., 1999. A noncalibrated rainfall—runoff model for large, arid catchments. Water Resources Research 35 (7), 2161—2172.

Lebel, T., Taupin, J.D., D'Amato, N., 1997. Rainfall monitoring during HAPEX Sahel 1. General rainfall conditions and climatology. Journal of Hydrology, 74—96.

McIntyre, N., Jackson, B., Wade, A.J., Butterfield, D., Wheater, H., 2005. Sensitivity analysis of a catchment-scale nitrogen model. Journal of Hydrology 315, 71—92.

McIntyre, N., Al-Qurashi, A., Wheater, H.S., 2007. Analysis of rainfall-runoff events from an arid catchment in Oman. Hydrological Sciences Journal 52 (6), 1103—1118.

Michaud, J.D., Sorooshian, S., 1994a. Effect of rainfall-sampling errors on simulations of desert flash floods. Water Resources Research 30 (10), 2765—2775.

Michaud, J.D., Sorooshian, S., 1994b. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. Water Resources Research 30 (3), 593—606.

Mott Macdonald, 1992. Groundwater recharge schemes for Saham-Sohar area — data analysis report. Sultanate of Oman, Ministry of Agriculture and Fisheries, Internal Report, May 1992, 48p.

Nouh, M., 2006. Wadi flow in the Arabian Gulf states. Hydrological Processes 20, 2393—2413.

Osborn, H.B., Renard, K.G., 1973. Thunderstorm runoff on the Walnut Gulch experimental watershed, Arizona, USA. Journal of Hydrological Division, ASCE Proceedings 99, 1129—1145.

Parrisopoulos, G.A., Wheater, H.S., 1992. Experimental and numerical infiltration studies in a wadi stream-bed. Hydrological Sciences Journal 37, 27—37.

Parlange, J.-Y., Rose, C.W., Sander, G., 1981. Kinematic flow approximation of runoff on a plane: an exact analytical solution. Journal of Hydrology 52 (1—2), 171—176.

Pilgrim, D.H., Chapman, T.G., Doran, D.G., 1988. Problems of rainfall—runoff modelling in arid and semiarid regions. Hydrological Sciences Journal 33 (4), 379—400.

Renard, K.G., Asce, M., Keppel, R.V., 1966. Hydrographs of ephemeral streams in the southwest. Journal of Hydraulics Division, ASCE Proceedings 2, 33—53.

Rodier, J.A., 1985. Aspects of arid zone hydrology. In: Rodda, J.C. (Ed.), Facets of Hydrology, vol. II. John Wiley and Sons Ltd., Chichester, pp. 205—247.

Saudi Arabian Dames and Moore, 1988. Representative basin study for wadis Yiba, Habawnah, Tabalah, Liyyah, and Lith, Report for Ministry of Agricultural and Water, Riyadh, Saudi Arabia.

Semmens, D.J., Goodrich, D.C., Unkrich, C.L., Smith, R.E., Woolhiser, D.A., Miller, S.N., 2008. KINEROS2 and the AGWA modeling framework. In: Wheater, H.S., Sorooshian, S., Sharma, K.D. (Eds.), Hydrological modelling in arid and semi-arid areas. Cambridge University Press, Cambridge, pp. 41—48.

Sharma, K.D., Murthy, J.S.R., 1996. Ephemeral flow modelling in arid regions. Journal of Arid Environments 33, 161—178.

Smith, R.E., Goodrich, D.C., Woolhiser, D.A., Unkrich, C.L., 1995. KINEROS — a kinematic runoff and erosion model. In: Singh, V.J. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, pp. 697—732.

Sorman, A.U., Abdulrazzaq, J., 1993. Infiltration-recharge through wadi beds in arid region. Hydrological Sciences Journal 38 (3), 173—186.

Walters, M.O., 1990. Transmission losses in arid region. Journal of Hydraulic Engineering, ASCE Proceedings 116 (1), 129—138.

Wheater, H.S., 1981. Northern Oman Flood Study. Imperial College, London, p. 270.

Wheater, H.S., 2002a. Progress in and prospects for fluvial flood modelling. Philosophical Transactions of the Royal Society of London Series A — Mathematical Physical and Engineering Sciences 360 (1796), 1409—1431.

Wheater, H.S., 2002b. Hydrological processes in arid and semi-arid areas. In: Wheater, H.S., Al-Weshah, R.A. (Eds.), Hydrology of Wadi Systems. IHP-V, No. 55. UNESCO, Paris.

Wheater, H.S., 2005. In: Modelling Hydrological Processes in Arid and Semi Arid Areas — an Introduction to the Workshop. GWADI Workshop, Roorkee. Chapter 1, Proceedings of the GWADI International Modelling Workshop, Roorkee, India, 39p. <http://www.gwadi.org/shortcourses>.

Wheater, H.S., Brown, R.P.C., 1989. Limitations of design hydrographs in arid areas — an illustration from south west Saudi Arabia. In: Proceedings of the 2nd British Hydrological Society National Symposium, Sheffield, September 1989, pp. 3.49—3.56.

Wheater, H.S., Butler, A.P., Stewart, E.J., Hamilton, G.S., 1991. A multivariate spatial—temporal model of rainfall in southwest Saudi Arabia. I. Spacial rainfall characteristics and model formulations. Journal of Hydrology 125, 175—199.

Wheater, H.S., Jakeman, A.J., Beven, K.J., 1993. Progress and directions in rainfall runoff modelling. In: Jakeman, A.J., Beck, M.B., McAleer, M.J. (Eds.), Modelling Change in Environmental Systems. John Wiley and Sons, Chichester, pp. 101—132.

Wheater, H.S., Chandler, R.E., Onof, C.J., Isham, V.S., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., Segond, M.L., 2005. Spatial—temporal rainfall modelling for flood risk estimation. Stochastic Environmental Research and Risk Assessment 19 (6), 403—416.

Woolhiser, D.A., Smith, R.E., Goodrich, D.C., 1990. KINEROS, A Kinematic Runoff and Erosion Model: Documentation and User Manual. US Department of Agriculture, Agricultural Research Service, ARS-77, 130p.

Yang, C., Chandler, R.E., Isham, V.S., Annoni, C., Wheater, H.S., 2005. Simulation and downscaling models for potential evaporation. Journal of Hydrology 302, 239—254.

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C.L., Goodrich, D.C., Schaffner, M., Stewart, A., 2008. Understanding uncertainty in distributed flash flood forecasting for semiarid regions. Water Resources Research, in press, doi:10.1029/2007WR005940.