New Approaches to Hydrological Prediction in Data-sparse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 333, 2009.

69

A probability distribution function approach to modelling rainfall–runoff response for data-sparse catchments

BINQUAN LI & ZHONGMIN LIANG

State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China
libinquan@gmail.com, zmliang@hhu.edu.cn

Abstract A rainfall–runoff (RR) model considering the spatial variation of rainfall, soil infiltration capability and soil storage capacity over a catchment and based on probability distribution functions, is used for rainfall–runoff modelling. The model combines infiltration excess (Horton) and saturation excess (Dunne) mechanisms. Moreover, it is applied to a data sparse catchment. Model parameters of the studied data sparse catchment are inferred from its parent gauged basin. In addition, a semi-distributed RR model called TOPMODEL is also employed in the parent gauged basin for comparison. Results show that the RR model can, to a certain extent, be applied to data sparse regions based upon hydrological similarity between the study catchment and its parent basin.

Key words spatial variation; probability density functions; rainfall–runoff models; data sparse regions; Yellow River

INTRODUCTION

Precipitation is a highly variable atmospheric variable and this makes it difficult to provide its accurate spatial and quantitative description. At present, raingauges are most widely used in rainfall estimation, but they cannot describe the spatial variability of rainfall (Niemczynowicz, 1986). Liang *et al.* (1996) considered that two approaches can be taken to model the spatial variability of precipitation. One is the pixel-based approach, which discretizes the precipitation over a spatial domain. Another option is the statistical dynamic approach to representing the spatial variability in precipitation. For the former approach, the determination of parameters is complicated and its data requirement is much stricter, while the statistical dynamic approach sometimes may produce better results. Consequently, the latter has been identified as a valuable hydrological descriptor and studied for many years. Warrilow *et al.* (1986) combined a negative exponential precipitation distribution with a constant maximum surface infiltration rate to estimate runoff from a catchment. In this paper, the precipitation probability density function (pdf) they presented is adopted. In addition, the spatial distribution of infiltration capacity and soil moisture storage capacity over a catchment are also non-uniform. Sometimes, they can also be characterized by probability distribution functions. The Xinanjiang model (Zhao, 1992), which has been successfully and widely applied in humid and semi-humid regions in China, expresses the spatial distribution of soil moisture in the form of a probability distribution function, similar to that advocated by Wood *et al.* (1992) and Todini (1996). In this paper, a simple RR model (Liang *et al.*, 2006) is employed. Based on the runoff-yield model, the quasi-analytical solution for surface runoff is deduced by the joint distribution of rainfall and soil infiltration capacity based on the Horton mechanism, while the underground runoff can be obtained according to Dunne mechanism.

However, model efficiency mainly depends on the hydro-meteorological data rather than the model itself. In many regions, the local data are often sparse or non-existent. For example, data collection is difficult in many montane areas, but these regions are important runoff source areas – generating floods, sustaining base flows and regulating the hydrological regime that maintains freshwater ecosystems (Bales *et al.*, 2006). Due to limited hydrological monitoring of these regions, a highly sophisticated hydrological modelling approach cannot be implemented (Kim & Kaluarachchi, 2008). In order to predict streamflow for data-limited catchments, model parameters could be extrapolated from gauged basins. In this paper, the RR model has a simple structure and few parameters. Moreover, its parameters are easy to determine, which means the model could be applied to data sparse basins. Several simple rainfall relationships are used as converters, and meanwhile the model parameters are determined based on hydrological similarity between the data

sparse catchment and its parent gauged basin. Consequently, the RR model is then tested in a data sparse catchment.

THE RAINFALL–RUNOFF MODEL

The structure of this RR model (Liang *et al.*, 2006) includes two main components. One is the surface runoff component: based on the Horton mechanism, the quasi-analytical expression of surface runoff can be deduced; the other is the underground runoff component: runoff generated below ground surface can be calculated according to the Dunne mechanism. Therefore, the total runoff is composed of Horton runoff and Dunne runoff. The model is presented in detail by Liang *et al.* (2006). However, a brief description is given here for completeness.

Surface runoff

In order to describe the spatial variability of rainfall within a basin, Warrilow *et al.* (1986) supposed that it could be reflected with a pdf which can be expressed as:

$$f(P_i) = \mu / P_m \cdot \exp(-\mu P_i / P_m) \text{ with } 0 \le \mu \le 1 \tag{1}$$

where P_i is rainfall intensity at any point within a catchment, P_m the catchment average rainfall intensity, and μ is a fraction representing the ratio of rainfall coverage area.

Here, the spatial variation of soil infiltration capacity is described by a parabola type cumulative distribution function (cdf):

$$F(F_i) = 1 - \left(1 - F_i / F_{mm}\right)^n \text{ with } F_{mm} = (n+1) F_m \tag{2}$$

where F_i is the surface infiltration rate, F_m the mean areal infiltration rate over a catchment, F_{mm} represents the maximum infiltration rate, and n is the exponent parameter measuring the non-uniformity of this distribution.

Surface runoff rate at point i is $P_i - F_i$ if $P_i \ge F_i$, or zero if not. For the former case, based on the randomness of surface runoff rate in space and the hypothesis that F_i and P_i are considered as two independent random variables for the sake of simplicity, the probability distribution of surface runoff rate, RS_i, can be obtained:

$$F(RS_i) = \iint\limits_{P_i - F_i < RS_i} f\left(F_i, P_i\right) dF_i dP_i = \exp(-P_{\min} / P_m) - nk_r / F_{mm} \cdot \exp(-\mu RS_i / P_m) \tag{3}$$

$$\text{with } k_r = \int_0^{F_{mm}} \exp(-\mu F_i / P_m) \cdot (1 - F_i / F_{mm})^{n-1} dF_i \tag{4}$$

where $P_{\min}$ is the minimum rainfall intensity while $P_{\max}$ the maximum rainfall intensity, and k_r can be calculated by a numerical integral approach. Moreover, the catchment average surface runoff, RS, which is the catchment average surface runoff rate multiplied by time step, Δt, can be expressed as:

$$RS = \int_0^{P_{\max}} RS_i dF(RS_i) \cdot \Delta t = nk_r P_m \Delta t / \mu F_{mm} \cdot \left[1 - \left(1 + \mu P_{\max} / P_m\right) \exp(-\mu P_{\max} / P_m)\right] \tag{5}$$

Consequently, the catchment average surface infiltration, F, is $F = P_m \Delta t - RS$.

Underground runoff

Infiltration is used for soil evaporation and to supplement soil moisture. The spatial distribution of soil moisture capacity is expressed as a probability distribution function (Zhao, 1992). It is demonstrated that the following relation holds reasonably well between the area at saturation, α, and the local proportion of maximum soil moisture content W' / W_{mm}, where W' is the point soil moisture at saturation and W_{mm} is the maximum possible soil moisture at any point over a basin.

$$\alpha = 1 - \left(1 - W' / W_{mm}\right)^b \text{ with } W_{mm} = (1+b)W_m \tag{6}$$

where W_m is the average soil moisture storage capacity, and b is an exponent parameter measuring the non-uniformity of this distribution. In addition, A is supposed as the point soil moisture at saturation corresponding with the average soil moisture content, W. Thus, A is defined as:

$$A = W_{mm}[1 - \left(1 - W/W_m\right)^{1/(b+1)}] \tag{7}$$

The underground runoff, RG, would be equal to zero if infiltration is less than the actual soil evaporation, E. And E is assumed to coincide with the potential evaporation which is equal to the reduction coefficient of evaporation, k, multiplied by the pan evaporation, E_w. Otherwise, there would be two possibilities:

(a) $RG = F - E - W_m + W + W_m\left[1 - \left(F - E + A\right)/W_{mm}\right]^{b+1}$, if $A + F - E < W_{mm}$; and

(b) $RG = F - E - \left(W_m - W\right)$, if not.

These equations, which represent the average underground runoff produced in the catchment, must be associated with an equation of state in order to update the mean water content in the soil. This equation takes the form:

$$W_{t+\Delta t} = W_t + F_t - E_t - RG_t \tag{8}$$

where the quantities represent averages over the catchment and their variation indicates changes between t and $t + \Delta t$, while W_t and $W_{t+\Delta t}$ are their values at time t and time $t + \Delta t$, respectively.

MODEL CALIBRATION AND VALIDATION

Dongwan catchment

Dongwan catchment, a sub-basin at the middle reach of the Yellow River in China, was studied. It is a humid and semi-humid area and it extends over about 2567.4 km^2. Its terrain is high in the west and low in the east. The mean annual precipitation is 744 mm. Rainstorms usually occur during July to October. Floods are mainly formed by rainstorms with high peak flow and short duration. There are 11 raingauges in the catchment (shown in Fig. 1). The figure also shows the geographical position and stations of Lushi catchment which is used as the case study in the discussion section.
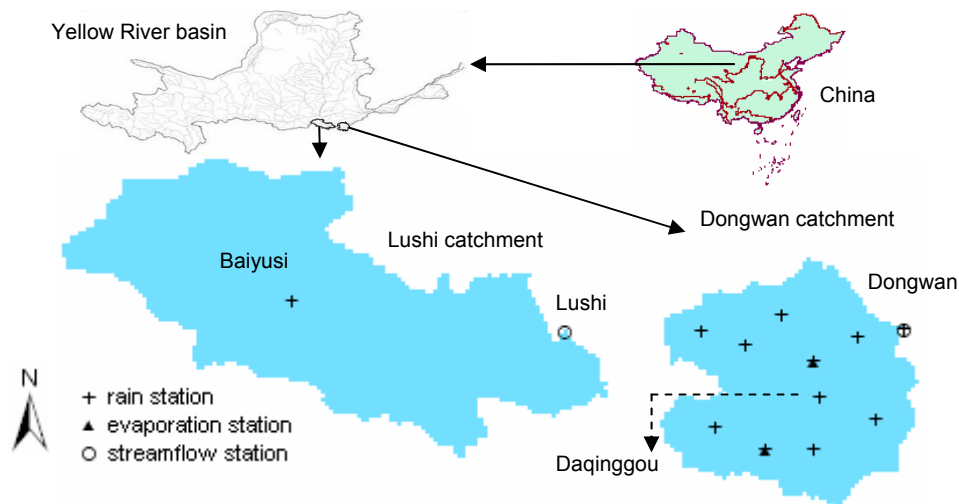

Fig. 1 Geographical position and stations of study catchments.

Model calibration and validation

In order to determine the mean areal infiltration rate over a catchment, the Horton infiltration model (Horton, 1940) is introduced here. The infiltration capacity, f_p, and time t relationship may be expressed as:

$$f_p = f_c + (f_0 - f_c) \cdot \exp(-\beta \cdot t) \tag{9}$$

where f_0 is the initial maximum infiltration rate, f_c is the final constant infiltration rate, and β is an exponent that controls the rate of decrease in the infiltration capacity. Based on equation (9), the mean areal infiltration capacity between t and $t + \Delta t$, $F_m = \int_t^{t+\Delta t} f_p \mathrm{d}t$, can be calculated. Thus the mean areal infiltration rate can be defined as:

$$F_m = f_c + 1/(\beta \cdot \Delta t) \cdot (f_0 - f_c) \exp(-\beta \cdot t)[1 - \exp(-\beta \Delta t)] \tag{10}$$

Based upon the category of soil and cover complex, the pre-set values of Horton model parameters can be determined according to Skaggs & Khaleel (1982). Table 1 gives all the parameters used in Dongwan catchment. Seven rainfall–runoff events of this basin are chosen for calibration and validation. Here, the time step is one hour. For every event, simulated runoff is compared with the corresponding observed result (Table 2) which is calculated by analysing component parts of its natural hydrograph. It can be seen that for all events, the absolute values of relative error are less than 20%, while the minimum value is 1.9% and the maximum is 17.7%. Given that rainfall for the seven rainfall–runoff events ranges from 19.0 to 170.9 mm, these rainfall–runoff events can be considered as representative. In addition, TOPMODEL (Beven & Kirkby, 1979) was applied to this basin, and its results are shown in Table 2 for comparison. It can be seen that the results obtained by the RR model are acceptable when compared with TOPMODEL, thus the RR model can be considered to be available for Dongwan catchment.

Table 1 Calibrated model parameter values in Dongwan catchment.

Parameter	μ	f_0 (mm/h)	f_c (mm/h)	β (h^{-1})	n	W_m (mm)	b	k
Value	1.0	670	20	84	0.4	130	0.4	0.9

Table 2 Calibration and validation results of the RR model and TOPMODEL in Dongwan catchment (hourly hydrological data of No.1-No.5 rainfall–runoff events are used for calibration; hourly hydrological data of No.6 and No.7 rainfall–runoff events are used for validation).

Rainfall–runoff event	Rainfall	Observed runoff	RR model Predictions	Error	TOPMODEL Predictions	Error
	(mm)	(mm)	(mm)	(%)	(mm)	(%)
(1)	(2)	(3)	(4)	[(4)–(3)]/(3)	(5)	[(5)–(3)]/(3)
1. 27 Jul.–9 Aug. 1996	170.9	105.1	123.7	17.7	117.9	12.2
2. 11–16 Aug. 1998	76.5	36.5	37.2	1.9	37.5	2.8
3. 5–8 Jul. 1999	19.0	5.9	5.2	–12.7	5.3	–10.4
4. 25 Jul.–1 Aug. 2001	103.3	19.9	19.0	–4.1	21.4	7.6
5. 17–18 Aug. 2005	60.8	13.6	12.7	–7.0	13.8	1.2
6. 18–20 Aug. 2005	32.2	11.2	11.6	3.5	11.6	3.1
7. 28 Sept.– 3 Oct. 2005	64.4	38.7	37.1	–4.0	36.3	–6.3

METHODOLOGY

Because there are many data-limited regions in the world, how to generate the flow series of these areas has been gaining more and more attention. In general, the following methods are commonly applied (Yu & Yang, 2000): (a) utilizing a model including only physically-based parameters that can be observed or inferred from measurements; or (b) extrapolating calibration parameters from those found at gauged sites close to the ungauged sites. The latter can also be interpreted as:

extrapolating model parameters calibrated by data of a gauged catchment to a data-limited catchment, thus, the model can be applied to such data sparse regions. Due to restrictions by many factors, hydrological models which have only physically-meaningful parameters are difficult to apply in actual predictions at present. Moreover, most RR models currently are described by some parameters which need to be determined from hydrological data. Consequently, it becomes very important to extrapolate the model parameters of data-rich regions to data-limited catchments.

Relationships between rainfall intensity of the central station and the catchment average/ maximum rainfall intensity of parent basins can be extrapolated to data sparse basins. For a rainfall–runoff event, first, we establish the relationship, for the parent gauged basin, between rainfall intensity of the central station, P_c, and the catchment average rainfall intensity based on the linear regression method, the relationship is called $P_c \sim P_m$ hereinafter; second, another relationship between rainfall intensity of the same central station and the catchment maximum rainfall intensity is also established which is similar to $P_c \sim P_m$ and called $P_c \sim P_{max}$ in this paper.

For the same event, the catchment average rainfall intensity of the data-limited basin at every time step can be obtained from the relationship $P_c \sim P_m$ of its parent basin according to P_c of the data-limited catchment. In a similar way, the catchment maximum rainfall intensity of the data-limited catchment can also be obtained from the relationship $P_c \sim P_{max}$. As a consequence, based on the extrapolated relationships, the catchment average/maximum rainfall intensity of the data sparse catchment can be inferred, thus the RR model can be applied to the data-limited catchment after all model parameters are determined.

CASE STUDY AND DISCUSSION

The study catchment

Lushi catchment, a sub-basin close to Dongwan catchment, is studied as a data sparse basin. Its area is about 4716 km^2, and its terrain is similar to Dongwan basin. However, there is only one raingauge, called Baiyusi, and one discharge station, called Lushi, which is the outlet, within the catchment. Details position of this catchment and its stations are given in Fig. 1.

Results and discussion

Based on the hydrological similarity, Dongwan catchment is chosen as the parent gauged basin of Lushi catchment in this section. One can see from Fig. 1 that Baiyusi station and Daqinggou station can be viewed as the central stations of their catchments, respectively. In addition, seven rainfall–runoff events used in Dongwan catchment are also used for this data sparse basin. The evaporation data of Dongwan catchment are directly extrapolated to Lushi catchment. Model parameters used in Lushi catchment are the same as those shown in Table 1, except for $n = 0.45$ on account of the difference of catchment area.

The simulated results are given in Table 3. The model efficiency of Dongwan catchment is better than that of Lushi catchment. This implies that the local data are even more important than all other influencing factors in rainfall–runoff modelling. It is also observed that the RR model can be used in Lushi catchment for rainfall–runoff modelling. For No.3 rainfall–runoff event, the absolute value of the runoff error is greater than 20%. This is probably because the hydrological data used in this rainfall–runoff event, which are inferred from Dongwan catchment, are under–representative of conditions in Lushi catchment.

As stated above, the RR model and the methodology presented are available for two study catchments. However, two main pre-conditions should be discussed: (a) model applicability, and (b) hydrological similarity. In order to verify the applicability of the RR model to the parent gauged basin, the pdf and cdf of equation (1) are compared with those of the statistical histogram of observed precipitation data for sufficient time steps different rainfall–runoff events. The comparison show that the RR model can be applied to Dongwan catchment. For instance, Fig. 2 gives the results of a certain time step, and this can validate the applicability to some extent.

Table 3 Validation of the RR model in Lushi catchment.

Rainfall–runoff event	Rainfall (mm)	Predictions (mm)	Observations (mm)	Error (%)
(1)	(2)	(3)	(4)	[(3)–(4)]/(4)
1. 27 Jul.–9 Aug. 1996	44.0	19.9	24.8	–19.9
2. 11–16 Aug. 1998	43.5	17.8	19.9	–10.3
3. 5– 8 Jul. 1999	20.7	5.8	4.8	21.4
4. 25 Jul.–1 Aug. 2001	31.0	2.9	3.4	–15.1
5. 17–18 Aug. 2005	21.6	3.1	3.6	–13.1
6. 18–20 Aug. 2005	20.4	6.8	6.3	8.0
7. 28 Sep.–3 Oct. 2005	72.4	47.4	49.8	–4.7
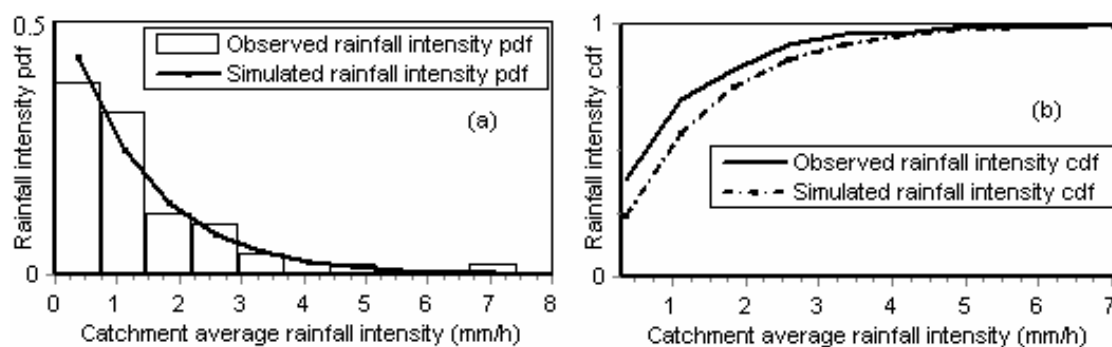
Fig. 2 Comparison of results between 19:00h 4 August 1996 and 20:00h 4 August 1996 of No.1 rainfall–runoff event: (a) pdf; and (b) cdf.

CONCLUSIONS

This paper introduces a simplified RR model coupling Horton and Dunne mechanisms, which is based upon the probability distribution function approach. The approach is tested in a humid and semi-humid region which is a sub-catchment of the Yellow River basin, and compared with TOPMODEL that does not use probability distribution functions. The methodology which shows how to simulate rainfall–runoff response for data sparse catchments, is presented. In the frame of this methodology, the RR model can be applied to data sparse regions. Based upon the hydrological similarity between the two study basins, the catchment average and maximum rainfall intensity of Lushi catchment are inferred from that of Dongwan catchment. And the pre-conditions are also discussed. Results obtained from the case study show that the RR model methodology presented in this paper to simulate rainfall–runoff response for data sparse basins is reliable. The major conclusions from this study are summarized below.

For prediction in data sparse regions, the specialty of this study is that several probability distribution functions are used for rainfall–runoff modelling, i.e. the RR model, and the model parameters of data sparse catchments can be inferred from their parent gauged basins based upon the hydrological similarity and the statistical properties of the probability distribution function approach. The RR model uses probability distribution functions to describe the spatial distribution of the three main factors influencing rainfall–runoff. Its surface runoff and underground runoff are based on Horton and Dunne runoff mechanisms, respectively. Therefore, it is a generic RR model which can be applied to different regions. One can modify the parameter values of the Horton runoff module and/or Dunne runoff module so that the model is suitable for the actual runoff generation mechanism of study areas. As a consequence, the RR model can be applied to data sparse catchments due to the statistical properties and the commonality of the probability distribution function approach.

Future work will also verify the methodology of application of this RR model to data sparse regions. And different forms of rainfall intensity pdf maybe combined into the RR model in order

to represent different spatial distributions of rainfall intensity and improve model efficiency, thus the model applicability in data sparse regions would be effectively improved.

REFERENCES

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R. & Dozier, J. (2006) Mountain hydrology of the western United States. *Water Resour. Res*. **42**, W08432.

Beven, K. J. & Kirkby, M. J. (1979) A physically-based variable contributing area model of basin hydrology. *Hydrol. Sci. J*. **24**(1), 43–69.

Horton, R. E. (1940) An approach toward a physical interpretation of infiltration-capacity. *Soil Sci. Soc. Am. Proc*. **5**, 399–417.

Kim, U. & Kaluarachchi, J. J. (2008) Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia. *J. Hydrol*. **362**, 39–56.

Liang, X., Lettenmaier D. P. & Wood E. F. (1996) One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. *J. Geophys. Res.* **101**(D16), 21403–21422.

Liang, Z. M., Lu, B. & Jia, L. J. (2006) A runoff-yield model considering spatial variation of rainfall. In: *Proceedings of the Sino-Dutch Water Innovation Seminar (Chinese Part), English Version* (ed. by Ministry of Water Resources, P. R. China & Ministry of Transport, Public Works and Water Management, The Netherlands), 188–195. Shanghai, China.

Niemczynowicz, J. (1986) The dynamic calibration of tipping-bucket raingauges. *Nordic Hydrol*. **17**, 203–214.

Skaggs, R. W. & Khaleel, R. (1982) Infiltration. In: *Hydrologic Modeling of Small Watersheds,* (ed. by C. T. Haan, H. P. Johnson & D. L. Brakensiek), 4–166. ASAE Monograph no. 5, American Society of Agricultural Engineers, Michigan, USA.

Todini, E. (1996) The ARNO rainfall–runoff model. *J. Hydrol.* **175**, 339–382.

Warrilow, D. A., Sangster, A. B. & Slingo, A. (1986) Modelling of land-surface processes and their influence on European Climate. *Dyn. Clim. Tech. Note 38*, Met. Off.ice, Bracknell, Berkshire, UK.

Wood, E. F., Lettenmaier, D. P. & Zartarian, V. G. (1992) A land-surface hydrology parameterization with subgrid variability for general circulation models. *J. Geophys. Res*. **97**, 2717–2728.

Yu, P. S. & Yang T. C. (2000) Using synthetic flow duration curves for rainfall–runoff model calibration at ungauged sites. *Hydrol. Processes* **14**,117–133.

Zhao, R. J. (1992) The Xinanjiang model applied in China. *J. Hydrol.* **135**, 371–381.