A statistically-based runoff-yield model

ZHONGMIN LIANG, JUN WANG, YANG LEI & YE SHI

State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

zmliang@hhu.edu.cn

Abstract This paper describes a statistical-based rainfall–runoff model considering spatial variation of rainfall and other factors that control the runoff-yield process. Both infiltration excess and saturation excess runoff-yield mechanisms are coupled in this model. The spatial variation of a rainfall event is described by introducing a probability density function (pdf) constructed by combining a simple linear function with a negative exponential function. Meanwhile, the soil infiltration capacity over a basin is simulated by a parabolic type pdf. According to the joint probability distribution of rainfall and soil infiltration capacity, the probability distribution of the surface runoff can be derived, a quasi-analytical solution for surface runoff is obtained, and furthermore the surface runoff and soil infiltration are computed by adopting the infiltration excess mechanism. Part of the infiltration supplements the soil moisture, and the rest recharges to subsurface and groundwater flow through using a water storage capacity curve. The total runoff is composed of infiltration excess and saturation excess runoff. As an example, the model was applied to Lushi basin, a sub-basin at the middle reach of the Yellow River. Results were compared with those obtained by TOPMODEL; it is shown that the model presented in this study is applicable to such a basin with semi-humid or semi-arid hydrological characteristics.

Key words statistically-based runoff-yield model; spatial variation of rainfall; infiltration excess mechanism; saturation excess mechanism; TOPMODEL

INTRODUCTION

Hydrological models are useful tools to study the mechanism of rainfall–runoff processes. Based on runoff-yield mechanisms, hydrological models are categorized into infiltration excess and saturation excess runoff-yield mechanisms. The former is typically applicable in arid or semi-arid areas, while the latter is applicable in humid or semi-humid areas. The conceptual model integrating with the two mechanisms was developed in the 1990s. Sivapalan *et al.* (1987) provided a theoretical framework considering both infiltration excess and saturation excess runoff-yield mechanisms. These integrated models are expected to be used in semi-humid and semi-arid areas. Based on whether they consider runoff-yield spatial variation or not, hydrological models are classified into distributed semi-distributed and lumped models. With the development of RS and GIS techniques, distributed models with physical explanation were studied and quickly developed. In the process of runoff yield, these models take the inhomogeneous distribution of runoff-yield factors into account, and model parameters with theoretical basis can be explicitly and physically explained. Between the lumped and distributed models, is the semi-distributed model, among which TOPMODEL (Beven, 1979) is the most representative. However, the determination of parameters in distributed or semi-distributed models is rather complicated and the requirement for data is much stricter, thus limiting the applications of these models.

In order to simplify the model structure and keep the accuracy, a statistical approach is sometimes applied to describe the spatial variation of runoff-yield. Two major points should be considered for the runoff-yield influencing factors (Xu *et al.*, 1996). The first type of factor is the spatial variability of the underlying surface, including geology, topography and soil structure. Another factor is certain hydrological and meteorological conditions, such as rainfall and evaporation. Among these factors, the spatial variation of rainfall is important in influencing the soil moisture and runoff-yield. Warrilow *et al.* (1986) used a negative exponential-type function to express the distribution of the rainfall over a basin. Sivapalan & Wood (1986) studied the effect of spatial heterogeneity of soil and rainfall; they derived quasi-analytical solutions for the ponding time and infiltration rate. These studies demonstrated that using a probability density function (pdf) could offer a valuable alternative for characterizing the spatial heterogeneity of a rainfall event.

In this paper, a new rainfall pdf constructed by combining a simple linear function with a negative exponential function is put forward to reflect the rainfall spatial distribution. According to the joint spatial probability distribution of rainfall and soil infiltration, the empirical Horton model (Horton, 1933) was adopted to separate surface runoff and infiltration. Part of the infiltration water supplements the soil water, and the rest is taken as the subsurface and groundwater flow, which is computed by a water storage capacity curve (Zhao *et al.*, 1980) based on saturation excess mechanism. Therefore, the total runoff is composed of infiltration excess and saturation excess runoff. It also means that this model can be applied to areas where runoff-yield is controlled by both infiltration and saturation excess runoff-yield mechanisms.

As an example, the statistically-based model was applied to simulate floods for Lushi basin. In addition, results were compared with those of TOPMODEL. It indicated that the simulation accuracy of the model is acceptable, which may make the model applicable to such semi-arid and semi-humid areas.

MODEL DESCRIPTIONS

Model structure

The runoff of this statistical-based model is divided into two parts: the surface runoff and the subsurface and groundwater runoff. The former is calculated by the infiltration excess runoff-yield mechanism through using the joint distribution of rainfall and infiltration capacity. The latter is calculated by the saturation excess runoff-yield mechanism through using a water storage capacity curve.

Surface runoff

The spatial variation of rainfall in any time step of a rainfall event can be described by probability distribution functions or probability density functions (pdf). In this study, the net rainfall at any point within a basin is supposed to comply with:

$$f(P_i) = \frac{1}{m^2} P_i \exp\left(-\frac{P_i}{m}\right) \tag{1}$$

where, P_i is the net rainfall at point i in the basin; m is a parameter which can be determined by the rainfall data from all the gauges in the basin by using goodness-of-fit. Figure 1 illustrates an example of the pdf from equation (1).

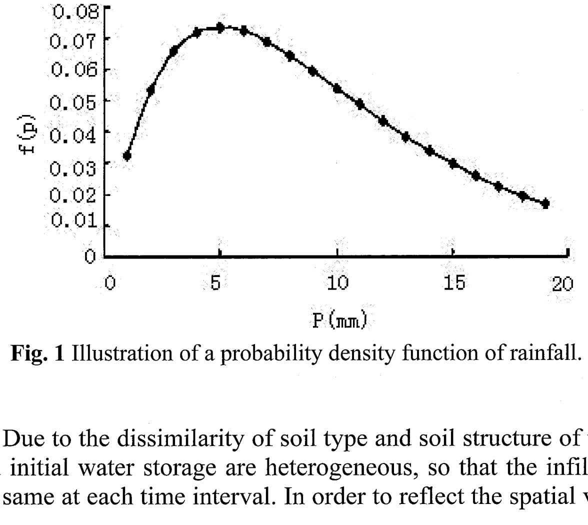

Fig. 1 Illustration of a probability density function of rainfall.

Due to the dissimilarity of soil type and soil structure of the basin, the water retention capacity and initial water storage are heterogeneous, so that the infiltration capability of each point is not the same at each time interval. In order to reflect the spatial variation of soil infiltration capacity, a cumulative distribution function (cdf) is introduced as:

$$F(F_i) = 1 - (1 - \frac{F_i}{F_{mm}})^n \tag{2}$$

where F_i is the soil infiltration capacity at point i. $F_{mm} = (n+1)F_m$ is the maximum soil infiltration capacity of the basin, while F_m is the mean of F_i. n is an index which reflects the unevenness of infiltration distribution. Based on equation (2), the pdf of soil infiltration capacity can be described as:

$$f(F_i) = \frac{dF}{dt} = \frac{n}{F_{mm}}(1 - \frac{F_i}{F_{mm}})^{n-1} \tag{3}$$

If $P_i > F_i$, the surface runoff generated at point i is $P_i - F_i$, and the rest infiltrates to the soil to supplement soil moisture or join in the generation of subsurface and groundwater flow. If $P_i \leq F_i$, all rainfall infiltrates to the soil moisture. As both the rainfall and infiltration capacity are treated as random, the surface runoff is also a random variable, therefore the cdf of surface runoff $R_s = P - F$ can be expressed as:

$$F(R_{si}) = \iint_{P_i - F_i < R_{si}} f(F_i, P_i) dF_i dP_i \tag{4}$$

where $f(F_i,P_i)$ is the joint distribution density function of dualistic variables (F_i,P_i). R_{si} is the surface runoff at point i. Soil infiltration capacity F_i and rainfall P_i are considered as two independent stochastic variables, thus equation (4) can be converted to:

$$F(R_{si}) = \int_0^{F_{mm}} \left[f(F_i) \int_{P_{\min}}^{R_{si}+F_i} f(P_i) dP_i \right] dF_i \tag{5}$$

where P_{min} is the minimum precipitation. Substituting equations (1) and (2) into equation (5):

$$F(R_{si}) = \int_0^{F_{mm}} \left[\frac{n}{F_{mm}}(1 - \frac{F_i}{F_{mm}})^{n-1} \int_{P_{\min}}^{R_{si}+F_i} \frac{1}{m^2} P_i \exp(-\frac{P_i}{m}) dP_i \right] dF_i \tag{6}$$

Solving the above integral, the distribution function of R_{si} can be deduced as:

$$F(R_{si}) = \frac{n}{F_{mm}} \left[k - \frac{k_{r1}}{m} \cdot R_{si} \cdot \exp(-\frac{P_{\max}}{m}) - (k_{r1} + \frac{1}{m} k_{r2}) \exp(-\frac{R_{si}}{m}) \right] \tag{7}$$

$$k = \frac{F_{mm}}{n} \cdot (\frac{1}{m} P_{\min} + 1) \cdot \exp(-\frac{P_{\min}}{m}) \tag{8}$$

$$k = \frac{F_{mm}}{n} \cdot (\frac{1}{m} P_{\min} + 1) \cdot \exp(-\frac{P_{\min}}{m}) \tag{9}$$

$$k_{r1} = \int_0^{F_{mm}} \exp(-\frac{F_i}{m}) \cdot (1 - \frac{F_i}{F_{mm}})^{n-1} dF_i \tag{10}$$

where k_{r1} and k_{r2} are temporary variables having no relation to R_{si}, and can be solved by numerical integration. So that the computation of surface runoff is essentially a quasi-analytical solution, the average or mean surface runoff can be calculated by:

$$R_s = E(R_{si}) = \int_0^{+\infty} R_{si} dF(R_{si}) \tag{11}$$

Generally, the upper limit of the integral is the maximum precipitation $P_{\max}$ of the rainfall event at the computation time step, then R_s can be computed as:

$$E(R_s) = \frac{n}{F_{mm} m}(k_{r2} + 2k_{r1} m) \left[m - (P_{\max} + m) \exp(-\frac{P_{\max}}{m}) \right] - \frac{nk_{r1}}{F_{mm} m} \cdot P_{\max}{}^2 \cdot \exp(-\frac{P_{\max}}{m}) \tag{12}$$

After the computation of surface runoff, the soil infiltration F can be computed by deducing surface runoff from rainfall P.

Subsurface and groundwater flow

When precipitation reaches the ground, some of it converts to surface runoff R_s, which can be calculated by equation (12), while the rest of it infiltrates to the soil. Part of the infiltration is consumed as soil evaporation, and part of it supplements soil moisture. Here we assume, at certain points within the basin, that subsurface and groundwater flow will occur when the soil water storage achieves field water capacity.

In order to reflect the unevenness of the infiltration capacity, a water storage capacity curve was brought in (shown as Fig. 2). From the infiltration and the curve, the subsurface and groundwater flow can be calculated. An empirical parabola function was adopted to compute runoff-yield area:

$$\alpha = 1 - (1 - \frac{W}{W_{mm}})^b \tag{13}$$

where, α is the ratio of runoff-yield area to the total basin area; W represents the water storage of a point within the basin; $W_{mm} = (1+b)W_m$ is the maximum water capacity of all points, while W_m is the mean areal water capacity; b is an index which reflects the uneven distribution of water storage capacity over the basin.

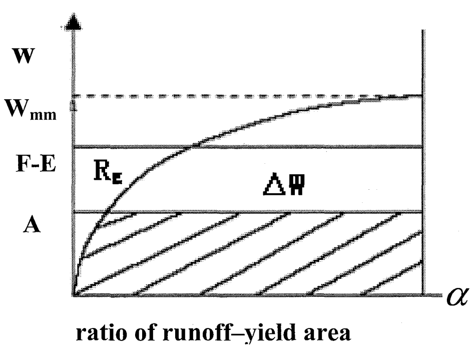

Fig. 2 Water storage capacity curve.

According to the initial mean areal water storage W_0, the corresponding point water storage capacity A (as shown in Fig. 2) can be computed as:

$$A = W_{mm}\left[1 - (1 - \frac{W_0}{W_m})^{\frac{1}{b+1}}\right] \tag{14}$$

The initial mean areal water storage W_0 at each time step is valued between 0 and W_m, which can be computed based on the water balance law.

$$W_0^t = W_0^{t-1} + F^{t-1} - E^{t-1} - R^{t-1} \tag{15}$$

where, W_0^t, W_0^{t-1} are denoted as mean areal water storage at time t and $t-1$ respectively; F^{t-1} is the soil water infiltration, while E^{t-1} is soil evaporation and R^{t-1} just represents the saturated runoff generated at time $t-1$. In addition, a coefficient k is introduced to compute actual evaporation E by $E = kE_p$, in which, E_p represents the soil evaporation potential.

If $F - E < 0$, the infiltration is used up for evaporation and no runoff, while $F - E > 0$, runoff is generated in following way:

when,

$$F - E + A < W_{mm} \quad R_g = F - E - W_m + W_0 + W_m[1-(F - E + A)/W_{mm}]^{1+b} \tag{16}$$

$$F - E + A \geq W_{mm} \quad R_g = F - E - (W_m - W_0) \tag{17}$$

R_g is denoted as subsurface and groundwater runoff. When the water storage of all points in the basin reaches their field water capacities, the total infiltrated water will convert to subsurface and groundwater runoff after the deduction of evaporation.

Parameter estimation

The model parameters can be classified into three groups: parameters of rainfall pdf, parameters of infiltration capacity pdf, and those of the water storage capacity curve. Parameter m can be determined by the goodness-of-fit approach. The mean areal maximum infiltration capacity F_m in the infiltration capacity distribution function can be computed by an empirical Horton model. Finally, index n, mean areal maximum water storage W_m, and the parabola index of water storage capacity curve b are normally determined through calibration.

APPLICATION

Study area

The Lushi basin, a sub-basin at the middle reach of Yellow River, is studied in this paper. Figure 3 is a draft map showing the location of Lushi basin and rainfall gauges within it. The control area of it is about 4716 km^2 and the mainstream length is about 180 km.

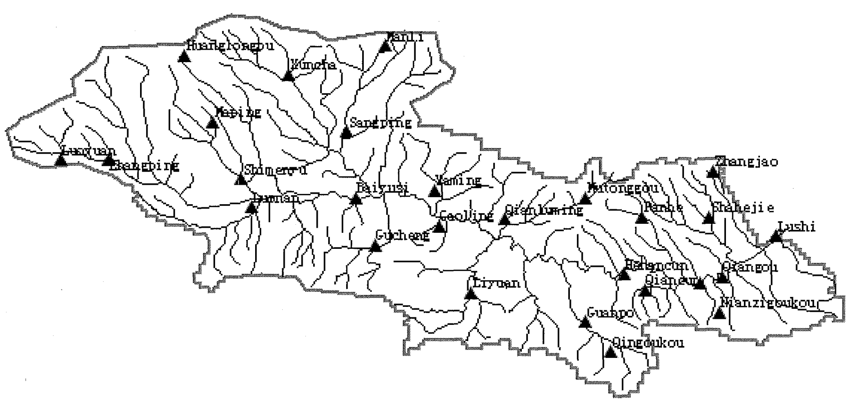

Fig. 3 Location of the rainfall gauges within Lushi basin.

The topographical elevation is higher in the west and lower in the east. The mean annual precipitation is around 600–800 mm. Rainstorms usually occur in July or August, and rainfall amount in these two months takes up 38% of the total yearly amount. Floods are mainly formed by storms, with characteristics of high peak flow and short duration.

Data

DEM data In this study, DEM data with resolution of 30″ × 30″ provided by USGS.

Rainfall gauges 26 rainfall gauges in the basin were used, which are shown in Fig. 3.

Evaporation Daily evaporation data at Lushi gauge were used to represent the aerial evaporation of the study basin.

Storm data Six storm events were selected with rainfall data at the 26 rainfall gauges and corresponding discharge data at Lushi gauge. The time step of each of the rainfall–runoff events is 1 hour.

Model application

The model runs for each of the time steps (1 hour) of a rainfall event, thus obtaining the simulated runoff process. The parameter m in the pdf of rainfall is estimated by rainfall data through using the goodness-of-fit method, and varies for each calculation time interval. n is an index to reflect the uneven feature of soil infiltration capacity over the basin and is fixed to 0.3 in this study. Generally, in the mountainous area, b is about 0.1 in small basins, 0.2–0.3 in medium scale basins and 0.3–0.4 in large basins. Here, b is given by 0.25. W_m is empirically defined to 130 mm. The model was used to simulate six flood events and results were compared with that of TOPMODEL.

RESULT ANALYSIS

Relative errors of the total runoff (in mm) for the six simulated flood events are listed in Table 1 for both the statistical-based runoff-yield model and TOPMODEL. For the statistical-based

Table 1 Comparison of observed and computed runoff depths.

				Statistical-based model		TOPMODEL	
Flood number	Start time	Mean areal precipitation (mm)	Observed runoff depth (mm)	Simulated runoff depth (mm)	Relative error (%)	Simulated runoff depth (mm)	Relative error (%)
1	1980070110	57.7	20.2	18.5	9	22.9	–13
2	1983072810	63.3	26.5	22.0	13	23.5	11
3	1984070410	72.1	23.2	23.7	–2	23.5	–1
4	1985091410	74.0	34.1	36.2	–6	29.8	12
5	1989070910	71.8	28.1	22.7	19	34.1	–21
6	1996080210	14.0	4.8	4.6	6	4.4	9

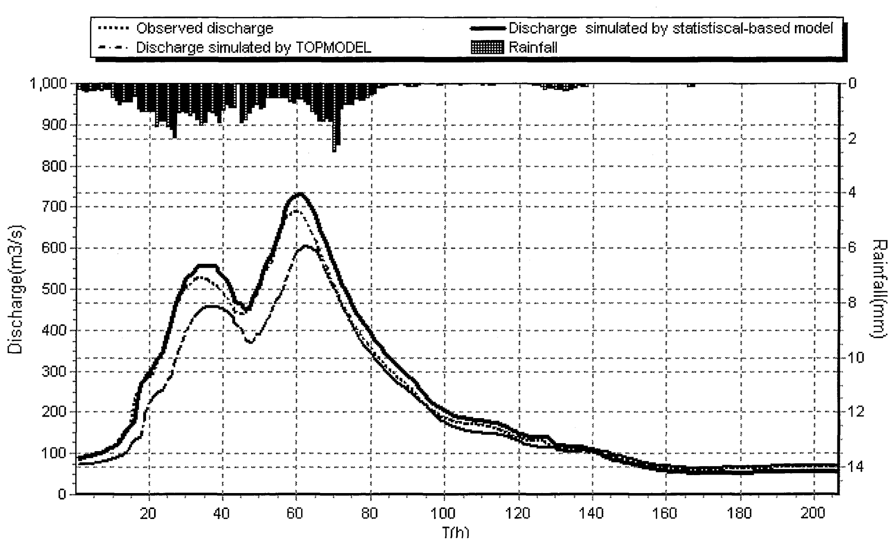

Fig. 4 Simulated and observed hydrographs of a flood event (flood # 1985091410).

model, the absolute value of the relative error for each flood event is less than 20%, with a minimum value of 2% and maximum value of 19%. For TOPMODEL, the absolute value of the relative errors for the six flood events are between 1 and 21%. As a whole it is indicated that the model is comparable to TOPMODEL, and could provide better simulations for rainfall–runoff events for the studied cases. As an example, the simulated and observed hydrographs of the flood event with start time 1985091410 is shown in Fig. 4.

SUMMARY AND CONCLUSION

A statistical-based runoff-yield model coupling the saturation and infiltration excess runoff-yield mechanisms is studied. A new pdf is introduced to describe the spatial distribution of rainfall, a quasi-analytical solution for runoff-yield calculation is obtained. The model was applied to simulate six flood events in Lushi basin, and results were compared with those obtained from TOPMODEL. It indicates that the model could achieve acceptable accuracy and is comparable to TOPMODEL for the studied flood events.

This statistically-based model considers the spatial variation of rainfall, soil infiltration capacity and water storage capacity that are the main factors controlling the runoff-yield process. Comparing the traditional approaches for calculation of areal rainfall such as arithmetic mean and Thiessen polygon, the application of a pdf possesses the advantage to describe the spatial heterogeneity of a rainfall event.

Acknowledgements This study was supported by the National Natural Science Foundation of China (grant no. 50779013), Specialized Research Fund for the Doctoral Program of Higher Education of China (grant no. 20070294018) and The National Basic Research Program of China ("973" Program, grant no. 2007CB7141973). We acknowledge two anonymous reviewers, whose suggestions have improved the paper.

REFERENCES

Beven K. J. & Kirkby, M. J. (1979) A physically based variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **27**(1), 43–69.

Horton, R. E. (1933) The role of infiltration in the hydrologic cycle. *Eos (AGU)* **14**, 446–460.

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**(16), 403–422.

Sivapalan, M. & Wood, E. F. (1986) Spatial heterogeneity and scale in the infiltration response of catchments. In: *Scale Problems in Hydrology* (ed. by V. K. Gupta, I. Rodriguez-Iturbe & E. F. Wood), 81–106. Reidel, Dordrecht, The Netherlands.

Sivapalan, M., Beven, K. & Wood, E. F. (1987) On hydrologic similarity 2. A scaled model of storm runoff production. *Water Resour. Res.* **23**(12), 2266–2278.

Warrilow, D. A., Sangster, A. B. & Slingo, A. (1986) Modeling of land-surface processes and their influence on European Climate. *Dyn. Clim. Tech.* Note **38**, 92 pp., Meteorological Office, Bracknell, Berkshire, UK.

Zhao, R. J., Zhuang, Y., Fang, L. R., Lin, X. R. & Zhang, Q. S. (1980) The Xinanjiang model. In: *Hydrological Forecasting to the Operation of Water Resource Systems* (Proc. Symposium, Oxford, April 1980), 351–356. IAHS Publ. 129, IAHS Press, Wallingford, UK.