Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Long-term probabilistic forecasting of snowmelt flood characteristics and the forecast uncertainty

LEV KUCHMENT & ALEXANDER GELFAN

Water Problems Institute of Russian Academy of Sciences, 119991 Gubkin 3, Moscow, Russia

kuchment@mail.ru

Abstract A methodology for long-term (i.e. with a lead time of 2–3 months) probabilistic forecasting of snowmelt flood volumes and peak discharges has been developed. The methodology is based on the use of a physically-based model of runoff generation combined with a weather generator. The distributed physically-based model includes description of snow accumulation and melt, soil freezing and redistribution of soil moisture during autumn and winter, and processes of runoff generation after the beginning of the spring snowmelt period. The weather generator consists of stochastic models of daily temperature and precipitation. The physically-based model has been applied to estimate missing initial river basin conditions before forecasting (usually, the soil moisture and depth of frozen soil; sometimes, the snow water equivalent), and the runoff hydrographs during the lead-time period using meteorological data for the autumn–winter period. The weather generator with Monte Carlo simulations, or selected weather scenarios, have been used to provide opportunities of assessing the meteorological inputs for lead-time periods and to estimate the probability distributions of the forecast runoff volumes and peak discharges. The results of forecasting of flood volumes have been compared with the results obtained on the basis of using the averaged meteorological conditions for lead-time periods and regression relationships between spring runoff volume and the initial indexes of river basin conditions before forecasting (the present day procedure for long-term flood forecasting). The case study was carried out for the Sosna River basin (catchment area 16 400 km^2) and for the Seim River basin (7500 km^2).

Key words long-term forecast; physically-based model; snowmelt flood; weather generator

INTRODUCTION

Long-term (with a lead time of 2–3 months) forecasting of spring–summer flood volume is widely used in Russia for water management and, especially, for runoff regulation of the large plain rivers where there are cascades of reservoirs. The traditional techniques of this forecasting are based on application of nonlinear or linear multiple regression relationships between flood volume and the main snowmelt runoff factors, which can include (depending on the physiographic region) snow water equivalent before melt, measured autumn soil moisture content (or its calculated index), and measured depth of frozen soil (or its calculated index). It is commonly assumed that the precipitation and the air temperature during the lead-time period are equal to their mean (climatological) values. However, these forecasting techniques can lead to significant errors if the conditions of snowmelt runoff generation are unusual (for example, the complicated processes of soil freezing and soil moisture redistribution, the high rate of snowmelt, several periods of thaw during a cold season, large liquid precipitation before melt, etc.), or meteorological conditions during the lead-time period significantly differ from climatological ones. We tried to improve the quality of the forecasts using continuous physically-based models of runoff generation driven by ensembles of meteorological variables during the lead-time period. As a result, we have obtained a technique that provides a probabilistic forecast of both snowmelt flood volume and flood peak discharge. The case study has been carried out for the Sosna River basin (catchment area 16 300 km^2) and for the Seim River basin (7500 km^2). Both basins are located in the forested-steppe zone of European Russia.

THE PHYSICALLY-BASED MODEL OF SNOWMELT FLOOD GENERATION

The model of runoff generation is based on the finite-element schematization of the river basin and describes the following main processes: snow cover formation and snowmelt, soil freezing and thawing, infiltration into frozen and unfrozen soil, vertical water transfer in an unfrozen soil, evaporation, overland and channel flow.

The system of vertically averaged equations of snow processes is applied to simulate temporal change of the snow depth, content of ice and liquid water, snow density, snowmelt, sublimation, re-freezing of melt water, snow metamorphism (Kuchment & Gelfan, 1996).

The melting rate S is determined as:

$$S = \beta \rho_s T_a \tag{1}$$

where ρ_s is the density of the snowpack; T_a is the air temperature; and β is the empirical constant.

The hydrothermal processes in the soil during the cold period are described by the following equations (Kuchment & Gelfan, 1993):

$$C_f \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(\lambda_f \frac{\partial T}{\partial z}\right) \quad \text{for } 0 < z < H(t) \tag{2}$$

$$C \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(\lambda \frac{\partial T}{\partial z}\right) \quad \text{for } H(t) < z < L \tag{3}$$

$$T(0,t) = T_0(t); T(H,t) = 0; T(L,t) = T_L; T(z,0) = T(z)$$

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left(D \frac{\partial \theta}{\partial z} - K\right) \quad \text{for } H(t) < z < L \tag{4}$$

$$\theta(L,t) = \theta_L; \theta(H) = \theta_0; \theta(\mathrm{z},0) = \theta(\mathrm{z})$$

$$\lambda_f \frac{\partial T}{\partial z}\Big|_{z=H-0} = \lambda \frac{\partial T}{\partial z}\Big|_{z=H+0} + \chi \rho_w (\theta_{-} - \theta_0) \frac{dH}{dt} \tag{5}$$

$$H(0) = 0$$

where $H(t)$ is the depth of frozen soil at time t; $T(z,t)$ is the soil temperature at the depth z and time t; λ_f and λ are the thermal conductivities of frozen and unfrozen layers of soil, respectively; C_f and C are the heat capacities of frozen and unfrozen layers of soil, respectively; χ is the latent heat of ice fusion; $\theta(z,t)$ is the volumetric liquid water content in the unfrozen layer of soil; $\theta\_$ is the liquid water content just above the freezing front; θ_0 is the liquid water content at 0°C (assumed to be equal to the moisture content at the wilting point); D is the diffusivity of soil moisture; K is the hydraulic conductivity of soil; L is the depth of the soil where the soil temperature and the volumetric moisture content can be considered as constants equal to T_L and θ_L, respectively (L was taken to be equal to 2 m).

Soil thawing is calculated for snow-free areas of the catchment area. The description of soil thawing model is presented in Kuchment *et al*. (2000).

Infiltration into a frozen soil is assumed to be equal to saturated hydraulic conductivity of the frozen soil K_f calculated as:

$$K_f = K_0\left(\frac{\theta_{\max} - I - \theta_0}{\theta_{\max} - \theta_0}\right)^4 \frac{1}{(1+8I)^2} \tag{6}$$

where K_0 is the saturated hydraulic conductivity of an unfrozen soil; $\theta_{\max}$ is the porosity; I is the ice content of the upper layer of soil.

The retention of water by the basin storage up to time t after the beginning of melting is determined as (Kuchment *et al*., 2000):

$$DET = DET_0\left[1 - \exp\left(-\frac{R}{DET_0}\right)\right] \tag{7}$$

where DET_0 is the mean value of the free storage capacity (or the maximum possible retention); R is the cumulated melt and rainfall water yield on the basin area up to time T. The changes of the unfrozen soil moisture content and infiltration into the soil during the warm period are calculated with the help of equation (4).

The evaporation rate E is calculated as (Kuchment & Gelfan, 1993):

$$E = k_E d_a \theta_u \tag{8}$$

where d_a is the air humidity deficit; θ_u is the soil moisture content at the upper soil layer; and k_E is the empirical constant.

To simulate overland and channel flow, one-dimensional kinematic wave equations are applied (Kuchment *et al*., 1986).

The runoff excess is calculated for each finite element taking into account the statistical distribution of snow water equivalent and the depth of frozen soil inside the element. The gamma distribution is applied for the both variables. The mean values are calculated by the model and coefficients of variation are derived from the empirical formulas (Kuchment & Gelfan, 1993).

Calibration and validation of the physically-based model

The soil hydraulic parameters in equations (4) and (6), and the soil thermal parameters in equations (2) and (3), were calculated by the empirical formulas presented in Kuchment *et al*. (1983). Soil properties which are necessary for using these formulas were taken from Hydromet. Survey (1975). Five parameters (the factor β, equation (1); the mean value of the free storage capacity DET_0, equation (7); the saturated hydraulic conductivity K_0, equation (6); and Manning's roughness coefficients for overland and channel flows) were calibrated against the measured daily river flow discharges. Calibration was carried out using data for the period of 1952–1966 for the Sosna River, and for the period 1947–1966 for the Seim River. The model validation was carried out using the measured discharges for 15 snowmelt floods (1967–1981) for the Sosna River and for 22 snowmelt floods (1967–1988) for the Seim River.

The efficiency criterion R^2 of Nash & Sutcliffe (1970) was adopted to estimate the goodness of fit of the simulated and measured characteristics of snowmelt floods. The results of the model validation are shown in Table 1.

Table 1 Results of the model validation.

	Number of floods used for the model validation	The efficiency criterion, R^2 Flood volume	Flood peak discharge
Sosna River	15	0.92	0.88
Seim River	22	0.90	0.88

THE WEATHER GENERATOR

The weather generator consists of the stochastic models of daily precipitation and air temperature for the period from 1 March to 30 April (for this period we neglect evaporation). The model of daily precipitation occurrence throughout this period is represented as the first-order Markov chain. Daily precipitation amount is described as a gamma-distributed random variable. The air temperature series are simulated by the procedure suggested in Kuchment *et al.* (2003). Parameters of the stochastic models were estimated using available meteorological records for the Sosna River basin for 30 years, and for the Seim River basin for 101 years.

FORECASTING TECHNIQUES AND RESULTS OF THEIR APPLICATION

Continuous simulation of runoff for each snowmelt flood began from 1 May of the previous year. Input data before the forecast issue date (1 March) included the observed daily precipitation, air temperature and air humidity. The initial conditions at the forecast issue date included the calculated characteristics of soil moisture and soil freezing, as well as the snow characteristics available from measurements or calculated (if measurements were not available). The input meteorological data for the lead-time period were constructed by one of the following three procedures:

1. Daily temperature and precipitation time series were assigned as the climatological means for each chronological day. In this case, it is possible to obtain the deterministic forecast of the spring flood hydrograph.
2. Daily temperature and precipitation time series were assigned as the available historical daily temperature and precipitation time series.

3. Daily temperature and precipitation time series were simulated using the weather generator and the Monte Carlo procedure.

Using the latter two procedures, we have an opportunity to calculate an ensemble of possible hydrographs, determine the probability distribution of the forecast flood volumes and peak discharges, and estimate the uncertainty of the forecast caused by different ways of assigning meteorological data sets constructed for the lead-time period.

The results of the deterministic forecasts of flood volume and peak discharge of 30 snowmelt floods (1952–1981) for the Sosna River, and 42 floods (1947–1988) for the Seim River are shown in Fig. 1.

These results were compared with the results from the forecast technique used in the present-day operational practice. For the forested-steppe zone of the European Russia, the present forecast technique is based on the empirical formula suggested by Apollov *et al.* (1974):

$$Y = (SWE + X_{cl}) - P_0\left[1 - \exp\left(-\frac{SWE + X_{cl}}{P_0}\right)\right] \tag{9}$$

$$P_0 = a\exp[-W(bH + c)]$$

where Y is the flood volume (mm); SWE is the snow water equivalent (mm); X_{cl} is the climatological mean of the precipitation amount for the lead time period; P_0 is the maximum possible runoff losses (mm); W is the index of moisture content of the 1 metre soil layer; H is the freezing depth (cm); and a, b and c are empirical coefficients (750 mm, 0.051 cm^{-1} and 0.11, respectively).

As can be seen from Table 2, the application of the physically-based model perceptibly improved the efficiency of the flood volume forecasts.

Table 2 Comparison of the deterministic forecasts of snowmelt flood volume based on the physically-based model and the traditional method (formula (9)).

	Number of the forecasted floods	The efficiency criterion R^2 Model	Formula (9)
Sosna River	30	0.81	0.72
Seim River	42	0.69	0.53

Using the second procedure of assigning meteorological data for the lead-time period, we calculated the ensemble of 30 hydrographs for each of the forecast floods at the Sosna River and the ensemble of 42 hydrographs for each of the forecast floods at the Seim River. Based on the third technique, we simulated 1000 time-series of meteorological inputs and calculated the ensemble of 1000 hydrographs for each of the forecast floods for both the rivers. The probability distributions of the forecast flood volume and peak discharge were obtained from the ensembles assigned by the both procedures.

To estimate the performance of the probabilistic forecasts obtained, we used two criteria:

1. Reduction of the original uncertainty of the flood characteristic Z due to the knowledge of the forecast z (transinformation $T(Z, z)$):

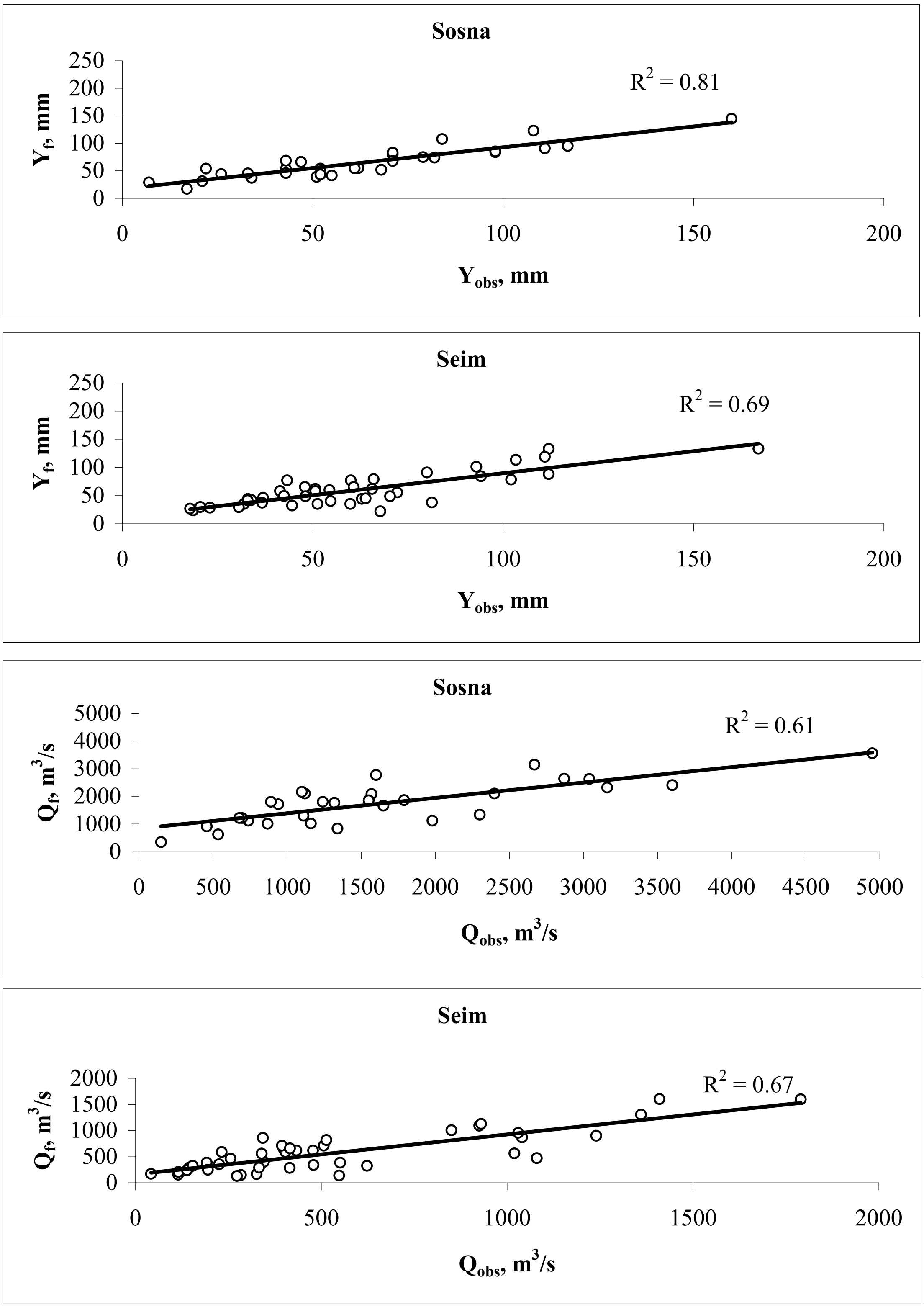

Fig. 1 Forecast *versus* observed snowmelt flood volumes (Y_f versus Y_{obs}) and peak discharges (Q_f vs Q_{obs}) for the Sosna and Seim rivers (deterministic forecast).

$$T(Z,z) = H(Z) - H(Z|z) \tag{10}$$

where $H(Z)$ is the marginal entropy of the observed flood characteristic Z; $H(Z|z)$ is the conditional entropy of Z given forecast z.

Accordingly to Amorocho & Espildora (1973), the marginal entropy $H(Z)$ and the conditional entropy $H(Z|z)$ can be calculated as:

$$H(Z) = \sum_{i=1}^{n} f(Z_i) \log \frac{1}{f(Z_i)} \Delta + \log \frac{1}{\Delta} \tag{11}$$

$$H(Z|z) = \sum_{i=1}^{n} \sum_{j=1}^{n} f(Z_i, z_j) \log \frac{1}{f(Z_i|z_j)} \Delta^2 + \log \frac{1}{\Delta} \tag{12}$$

where $f(Z)$, $f(Z,z)$, $f(Z|z)$ are the marginal, the joint and the conditional probability density functions, respectively, which we determined from the empirical histograms of the forecast and the observed flood characteristics; n is the number of intervals of width Δ which subdivide the total range of Z.

The ratio of transinformation $T(Z,z)$ to the original entropy $H(Z)$, i.e. $\delta = \dfrac{T(Z,z)}{H(Z)}$ can be considered as the relative estimate of the forecast efficiency.

2. Criterion *RPSS* (ranked probability skill score), which is used for verification of the probabilistic meteorological forecast (Wilks, 1995), is calculated as:

$$RPSS = 1 - \frac{\overline{RPS_f}}{\overline{RPS_{cl}}} \tag{13}$$

where $\overline{RPS}$ (ranked probability score) is the averaged sum of the squared deviations between the cumulative distributions of the forecasts and the observations; $\overline{RPS_f}$ is calculated using the forecasts by the model; and $\overline{RPS_{cl}}$ is calculated using the forecasts based on the historical distribution of observed values (climatology forecasts). A positive *RPSS* indicates improvement over climatology and that the forecasts provide additional predictive information.

Results of comparison of the two procedures considered of the probabilistic forecast are given in Table 3. As one can see from Table 3, both procedures give additional predictive information but the performance of the third forecast technique

Table 3 Performances of the probabilistic forecasts of flood characteristics obtained using two different procedures of assigning meteorological conditions for the lead time of the forecast.

Meteorological conditions used for the lead-time period	Relative reduction of original uncertainty, δ		Ranked probability skill score, *RPSS*	
	Flood volume	Flood peak discharge	Flood volume	Flood peak discharge
Sosna River				
Historical sets of meteorological conditions	0.24	0.15	0.66	0.53
Meteorological conditions simulated by the weather generator	0.26	0.17	0.73	0.57
Seim River				
Historical sets of meteorological conditions	0.19	0.15	0.61	0.55
Meteorological conditions simulated by the weather generator	0.22	0.21	0.63	0.59

(using the weather generator for the lead-time period) appeared to be slightly better than the second technique which is based on using historical series as the future weather. It is obvious that, in general, the difference of the results given by these procedures depends on the length of the available meteorological series and the quality of the weather generator; however, the weather generator can give a wider range of possible meteorological variables.

CONCLUSION

The suggested methodology of forecasting of snowmelt flood volumes and peak discharges using the physically-based model can be efficiently applied for both deterministic and probabilistic long-term flood forecasts. Probabilistic forecasts provide decision makers with additional information which can improve the usefulness of the flood forecast. However, it is necessary to investigate the benefit and reliability of this methodology for river basins in various physiographic zones and with different sets of available data. It is important to evaluate the dependence between the forecast quality and the lead time for each selected river basin.

Acknowledgement This research was supported by the Russian Foundation of the Basic Research (Grant 05-05-64828).

REFERENCES

Amorocho, J. & Espildora, B. (1973) Entropy in the assessment of uncertainty in hydrologic systems and models. *Water Resour. Res*. **9**(6), 1511–1522.

Apollov, B. A., Kalinin, G. P. & Komarov, V. D. (1974) *Hydrological Forecasting*. Gidrometeoizdat, Leningrad, USSR (in Russian).

Kuchment, L. S. & Gelfan, A. N. (1993) *Dynamic-Stochastic Models of River Runoff Generation*. Nauka, Moscow, USSR (in Russian).

Kuchment, L. S. & Gelfan, A. N. (1996) The determination of the snowmelt rate and meltwater outflow from a snowpack for modelling river runoff generation. *J. Hydrol*. **179**(1/4), 23–36.

Kuchment, L. S, Demidov, V. N. & Motovilov, Yu. G. (1983) *River Runoff Formation* (physically-based models). Nauka, Moscow, USSR (in Russian).

Kuchment, L. S., Demidov, V. N. & Motovilov, Yu. G. (1986) A physically-based model of the formation of snowmelt and rainfall runoff. In: *Symposium on the Modelling Snowmelt-Induced Processes* (ed. by E. M. Morris), 27–36. IAHS Publ.155. IAHS Press, Wallingford, UK.

Kuchment, L. S, Gelfan, A. N & Demidov, V. N. (2000) A distributed model of runoff generation in the permafrost regions. *J. Hydrol*. **240**(1/2), 1–22.

Kuchment, L. S., Gelfan, A. N. & Demidov, V. N. (2003) Application of dynamic-stochastic models of runoff generation for estimating extreme flood frequency distribution. In: *Water Resources Systems—Hydrological Risk, Management and Development* (ed. by G. Blöschl, S. Franks, M. Kumagai, K. Musiake & D. Rosbjerg), 107–114. IAHS Publ. 281. IAHS Press, Wallingford, UK.

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models. *J. Hydrol*. **10**(3), 282–290.

Hydromet. Survey (1975) *Reference Book of the Hydrological Properties of Soils in the Central-Chernozem Region.* Hydrometeorological Survey of the Central-Chernozem Regions, Kursk (in Russian).

Wilks, D. S. (1995) *Forecast Verification: Statistical Methods in the Atmospheric Sciences*. Academic Press, New York, USA.