*Weather Radar Information and Distributed Hydrological Modelling* (Proceedings of symposium HS03 held during IUGG2003 at Sapporo, July 2003). IAHS Publ. no. 282, 2003.

# Stochastic modelling of the error structure of real-time predicted rainfall and rainfall field generation

**YASUTO TACHIKAWA, YOSHIMITU KOMATSU, KAORU TAKARA**
*Disaster Prevention Research Institute, Kyoto University, Uji, Kyoto 611-0011, Japan*
tatikawa@mbox.kudpc.kyoto-u.ac.jp

**MICHIHARU SHIIBA**
*Department of Urban and Environmental Engineering, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan*

**Abstract** To evaluate the uncertainty of a real-time river discharge prediction with a distributed rainfall–runoff model, an error structure of real-time rainfall prediction by a translation model is modelled as a spatial random field, and then predicted rainfall fields are simulated according to the characteristics of the prediction error structure. At first, prediction error fields are statistically analysed and an isotropic lognormal spatial random field is selected to model relative prediction error fields. Next, a method to generate a lognormal spatial random field is developed, which uses a matrix factorization technique that a covariance matrix is decomposed into its square root matrix. Using this method, rainfall field realizations that take into account the prediction error structure are well simulated.

**Key words** precipitation forecasting; random field; translation model; uncertainty

## INTRODUCTION

For real-time rainfall and discharge prediction it is important to not only provide a forecasting value but also to give the prediction accuracy. So far, many methods of short-term rainfall prediction using weather radar data have been proposed, and some researchers have extended these methods to provide the uncertainty of real-time rainfall forecasting (Takasao *et al.*, 1994; Georgakakos & Krajewski, 1995; Georgakakos, 2000), in which a stochastic state-space form of a rainfall prediction model using operationally available radar data are presented. In this study, to evaluate the uncertainty of a river discharge prediction by using a physically based distributed hydrological model, the idea to solve a stochastic distributed hydrological system numerically on a Monte Calro simulation framework using rainfall prediction field simulations is adopted. If simulated alternative realizations of a predicted rainfall field having common statistical properties are provided as the input to a physically-based distributed hydrological model, the output discharges will produce the probability distribution of predicted discharges at any point in space. This is the objective of the research presented here.

In the next section, a radar data used here and a short-term rainfall prediction method using a translation model (Shiiba *et al.*, 1984; Takasao & Shiiba, 1985) are briefly described. After that, prediction error fields calculated with radar observed rainfall fields and predicted rainfall fields by the translation model are statistically analysed and it is found that the distribution of a relative prediction error fits to a lognormal distribution having a spatial correlation. Therefore an isotropic lognormal spatial random field model is selected to represent the spatial prediction error structure. Then a method to generate the lognormal spatial random field is developed, which uses a matrix factorization technique that a covariance matrix is decomposed into its square root matrix approximately by using the Chebyshev polynomials. Finally, rainfall field generations by using the method are demonstrated.

## RADAR RAINFALL DATA AND TRANSLATION MODEL

### Radar rainfall data

The radar data used are observed at the Miyama radar raingauge located at the centre of Kinki district, Japan, which is managed by the Ministry of Land, Infrastructure and Transport. The radar measures rainfall fields at 5-min intervals with the horizontal resolution of 3 km, covering the area of 120 km radius. In this study, the radar data of 80 × 80 grid cells (240 × 240 km with 3 km spatial resolution) observed between 5 September at 10:00 and 6 September at 6:00 in 1989 are used. By using the radar data, 12 kinds of averaged data with the combinations of time resolutions of 5, 15, 30, and 60 min and spatial resolutions of 3, 6 and 12 km are generated for statistical analyses.

### Translation model

The horizontal rainfall distribution $z(x, y, t)$ with the spatial coordinate $(x, y)$ at time $t$ is modelled as:

$$\frac{\partial z}{\partial t}+u\frac{\partial z}{\partial x}+v\frac{\partial z}{\partial y}=w \tag{1}$$

where $u$ and $v$ are advection velocity along $x$ and $y$, and $w$ is growth–decay rate. Here $u$, $v$ and $w$ are assumed to have the forms:

$$u=c_1x+c_2y+c_3,\quad v=c_4x+c_5y+c_6,\quad w=c_7x+c_8y+c_9 \tag{2}$$

The values of parameters $c_1$, ..., $c_9$ are identified by the least square method using observed radar rainfall data and they are updated on a real-time basis (Shiiba *et al.*, 1984; Takasao & Shiiba, 1985). Three consecutive spatial rainfall distributions in 15 min are used to determine $u$ and $v$, and rainfall fields with 5 min and 3 km resolution are predicted for each 5 min interval. After that, 12 kinds of averaged predicted data with different time and spatial resolution stated above are generated to analyse prediction error structures. When forecasting rainfall fields, the growth decay rate $w$ is always assumed to be zero.

## MODELLING OF REAL-TIME RAINFALL PREDICTION ERROR

### Model of prediction error field

The spatial distribution of an absolute prediction error $\boldsymbol{E}_a$ and the one of a relative prediction error $\boldsymbol{E}_r$ with the forms:

$$\boldsymbol{E}_a = \boldsymbol{R}_o - \boldsymbol{R}_p, \quad \boldsymbol{E}_r = (\boldsymbol{R}_o - \boldsymbol{R}_p) / \boldsymbol{R}_p \tag{3}$$

are calculated and statistically analysed, where $\boldsymbol{R}_o$ is an observed radar rainfall field and $\boldsymbol{R}_p$ is a predicted rainfall field by the translation model. The calculation in equation (3) is conducted for each corresponding grid cell of $\boldsymbol{R}_o$ and $\boldsymbol{R}_p$. As shown in equation (3), $\boldsymbol{E}_r$ cannot be obtained if $\boldsymbol{R}_p$ is zero, so $\boldsymbol{E}_r$ is assumed to be zero if $\boldsymbol{R}_p$ is zero.

If a random field model represents the spatial distributions of $\boldsymbol{E}_a$ and/or $\boldsymbol{E}_r$ and it is well simulated according to the statistical characteristics, a realization of rainfall field $\boldsymbol{R}$ is simulated with the equation:

$$\boldsymbol{R} = \boldsymbol{R}_p + \boldsymbol{E}_a \quad \text{or} \quad \boldsymbol{R} = \boldsymbol{R}_p \times \boldsymbol{E}_r + \boldsymbol{R}_p \tag{4}$$

The simulated alternative rainfall fields $\boldsymbol{R}$ are applied to a distributed rainfall–runoff model and the probability distribution of a discharge prediction will be obtained.

### Statistical characteristics of prediction error

Figure 1 shows the frequency distributions of an absolute prediction error $\boldsymbol{E}_a$ for 5-min ahead prediction, which are those calculated using grid-cells in which the predicted

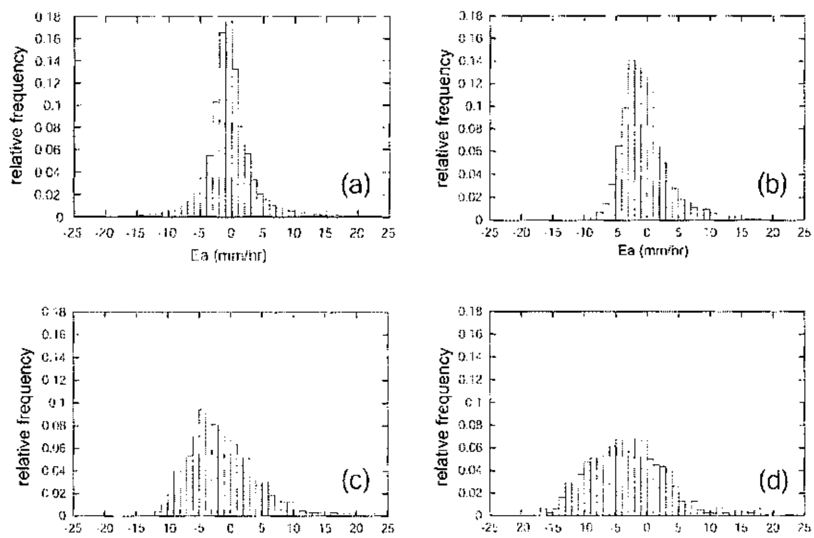


**Fig. 1** Frequency distribution of absolute prediction error $\boldsymbol{E}_a$ for 5-min ahead prediction: frequency distributions of $\boldsymbol{E}_a$ for grid cells in which predicted rainfall intensity is (a) larger than zero; (b) between 5 and 10 mm h$^{-1}$; (c) between 10 and 15 mm h$^{-1}$; and (d) between 15 and 20 mm h$^{-1}$.

rainfall intensity is larger than zero, between 5 and 10 mm h$^{-1}$, between 10 and 15 mm h$^{-1}$, and between 15 mm h$^{-1}$ and 20 mm h$^{-1}$, respectively. The shapes of frequency distributions are quite dependent on the predicted rainfall intensity and also $\boldsymbol{E_a}$ showed a positive spatial correlation. It is difficult to model a spatial random field having different probability distributions with a spatial correlation structure. Then, a relative prediction error $\boldsymbol{E_r}$ is tested. As shown in Fig. 2, every frequency distribution seems to fit to the same lognormal distribution with the lower bounded value, which takes –1 by equation (3) when the value of an observed rainfall is zero. By using the data, the spatial correlation coefficient of $\boldsymbol{E_r}$ is also calculated. In Fig. 3, it is found that $\boldsymbol{E_r}$ has a positive correlation within 10 km. From these results, $\boldsymbol{E_r}$ could be modelled as a lognormal random field with a spatial correlation structure.

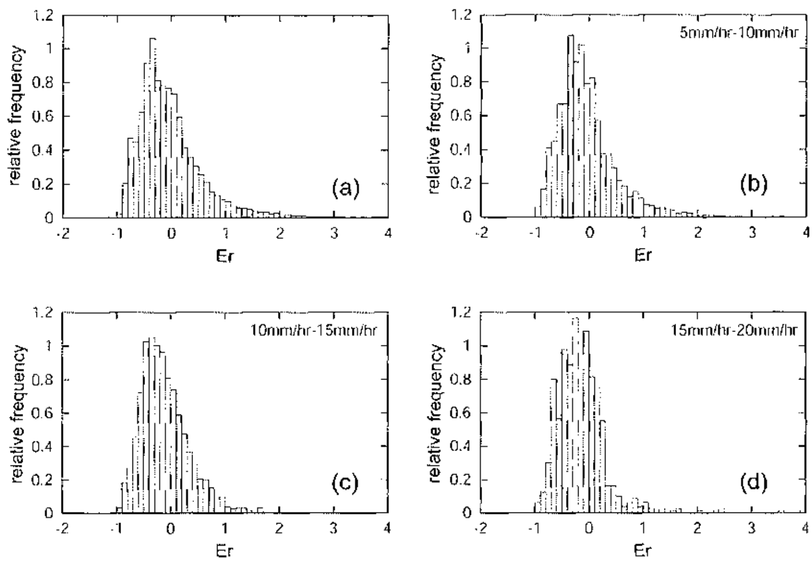


**Fig. 2** Frequency distribution of relative prediction error $\boldsymbol{E_r}$ of 5-min ahead prediction. Frequency distribution of $\boldsymbol{E_r}$ for grid cells in which predicted rainfall intensity is (a) larger than zero; (b) between 5 and 10 mm h$^{-1}$; (c) between 10 and 15 mm h$^{-1}$; and (d) between 15 and 20 mm h$^{-1}$.

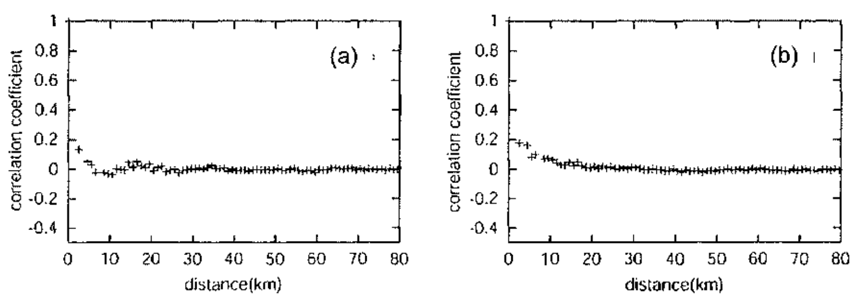


**Fig. 3** Spatial correlation coefficient of relative prediction error $\boldsymbol{E_r}$ (a) case for 5-min ahead prediction, and (b) case for 60-min ahead prediction.

**Table 1** Results of Kolmogorov-Smirnov-test of fit for relative prediction error to lognormal distribution. The significance level is 5%.

| Time resolution of radar rainfall | Spatial resolution of radar rainfall | 5-min ahead prediction | 15-min ahead prediction | 30-min ahead prediction | 45-min ahead prediction | 60-min ahead prediction |
|---|---|---|---|---|---|---|
| 5 min | 3 × 3 km | G | G | F | F | F |
| | 6 × 6 km | G | G | G | F | F |
| | 12 × 12 km | G | G | G | G | F |
| 15 min | 3 × 3 km | -- | G | F | F | F |
| | 6 × 6 km | -- | G | G | F | F |
| | 12 × 12 km | -- | G | G | G | G |
| 30 min | 3 × 3 km | -- | -- | F | -- | F |
| | 6 × 6 km | -- | -- | F | -- | F |
| | 12 × 12 km | -- | -- | G | -- | G |
| 60 mim | 3 × 3 km | -- | -- | -- | -- | G |
| | 6 × 6 km | -- | -- | -- | -- | G |
| | 12 × 12 km | -- | -- | -- | -- | G |

G: good fit; F: failed fit; --: not applicable

By using the Kolmogorov-Smirnov test, 12 kinds of relative prediction errors with different time and spatial resolutions calculated by averaged data stated above are tested to fit a lognormal distribution. Each relative prediction error result is classified into three categories according to predicted rainfall intensity having less than 4 mm $h^{-1}$, between 4 and 10 mm $h^{-1}$, and more than 10 mm $h^{-1}$, respectively. Then the frequency distribution for each classified data is tested to fit to the frequency distribution for all grid cells with rainfall intensity larger than zero with the criterion of 5% significance level of the Kolmogorov-Smirnov test. If all the frequency distributions of the three classified data fit to the frequency distribution of all grid-cells with rainfall, the distribution is supposed to represent the distribution of the relative prediction error. The results of the Kolmogorov-Smirnov test are summarized in Table 1. The analysis proposes to use spatially-averaged data with 12-km resolution or time averaged data with 1 h if a prediction lead time is taken to be 1 h.

## Characteristics of time persistency for relative precipitation error

To simulate a spatial field of $\boldsymbol{E_r}$ for future time, it is desirable that the statistical characteristics of $\boldsymbol{E_r}$ last for several hours. Figure 3 shows the mean and standard

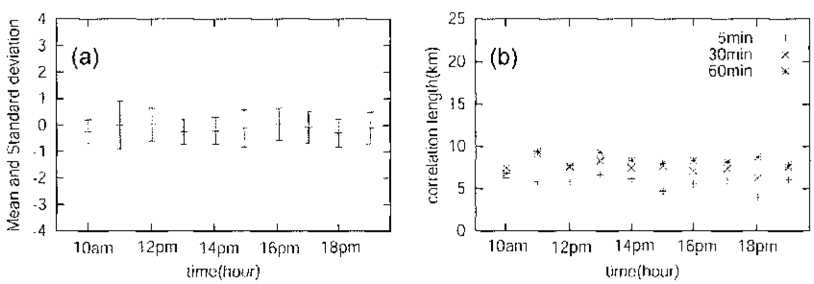


**Fig. 4** (a) Time variation of mean and standard deviation of relative prediction $\boldsymbol{E_r}$ for 60-min ahead prediction. (b) Time variation of correlation length $a$.

deviation of $\boldsymbol{E}_r$ calculated at every hour from 10:00 to 19:00 on 5 September 1989 by using the data with 3-km spatial and 1-h time resolution. It shows that the changes in mean and standard deviation are small during one precipitation duration. Figure 4 shows the change of spatial correlation length of $\boldsymbol{E}_r$, $\alpha$, the parameter of Gaussian spatial correlation function $\rho(h) = \exp(-h^2/\alpha^2)$, obtained by the least square method. The values of $\alpha$ are calculated every hour by using $\boldsymbol{E}_r$ of 5-, 30-, 60-min ahead prediction. The correlation length shows very slight changes in time for all cases. From the results, it is supposed that the statistical characteristics of $\boldsymbol{E}_r$ obtained at current time step will last several hours.

## METHOD FOR LOGNORMAL RANDOM FIELD SIMULATION

The method to use a matrix factorization technique is adopted in which a covariance matrix is decomposed into its square root matrix approximately (Davis, 1987; Tachikawa & Shiiba, 2000).

Let $\boldsymbol{R}$ be a given covariance matrix, and consider $\boldsymbol{R}$ to decompose into $\boldsymbol{R} = \boldsymbol{SS}$. Here, $\boldsymbol{S}$ is a symmetric positive-definite matrix defined as $\boldsymbol{S} = \boldsymbol{Q}\Lambda^{1/2}\boldsymbol{Q}^T$, where $\boldsymbol{Q}$ and $\Lambda^{1/2}$ are an orthogonal and a diagonal matrix defined by means of eigenvalue decomposition of $\boldsymbol{R}$ into $\boldsymbol{R} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^T$. Then a simulation generated by $\boldsymbol{X} = \boldsymbol{Sw}$ is a Gaussian random field where w is a vector of independent standard Gaussian random variables, because:

$$E[\boldsymbol{XX}^T] = E[\boldsymbol{Sww}^T\boldsymbol{S}^T] = \boldsymbol{Q}\,\Lambda^{1/2}\boldsymbol{Q}^T E[\boldsymbol{ww}^T]\,\boldsymbol{Q}\Lambda^{1/2}\boldsymbol{Q}^T = \boldsymbol{Q}\Lambda\boldsymbol{Q}^T = \boldsymbol{R}$$

To obtain $\boldsymbol{S}$, $\Lambda^{1/2}$ is numerically expanded by using the Chebyshev polynomials as:

$$\Lambda^{1/2} = \alpha_0\boldsymbol{I} + \alpha_1\Lambda + \alpha_2\Lambda^2 + \ldots + \alpha_p\Lambda^p + \Delta$$

and $\boldsymbol{S}$ is approximately obtained by:

$$\boldsymbol{S} = \boldsymbol{Q}(\alpha_0\boldsymbol{I} + \alpha_1\Lambda + \alpha_2\Lambda^2 + \ldots + \alpha_p\Lambda^p + \Delta)\boldsymbol{Q}^T =$$

$$\alpha_0\boldsymbol{I} + \alpha_1\boldsymbol{R} + \alpha_2\boldsymbol{R}^2 + \ldots + \alpha_p\boldsymbol{R}^p + \boldsymbol{Q}\Delta\boldsymbol{Q}^T \quad (5)$$

Here $\boldsymbol{R}^n = (\boldsymbol{Q}\Lambda\boldsymbol{Q}^T)^n = \boldsymbol{Q}\Lambda^n\boldsymbol{Q}^T$ is used to derive equation (5).

After getting a homogeneous Gaussian random field $\boldsymbol{X}$ with mean $m_X$, variance $\sigma_X^2$, and covariance function $C_X(h)$ using the above method, a lognormal random field $\boldsymbol{Y} = \exp[\boldsymbol{X}]$ with mean $m_Y$, variance $\sigma_Y^2$, and covariance function $C_Y(h)$ is finally obtained. The procedure for simulating a lognormal random field giving $m_Y$, $\sigma_Y^2$, and $C_Y(h)$ is as follows: (a) calculate $\sigma_X^2$ using the relationship $\sigma_Y^2 = m_Y^2[\exp\{\sigma_X^2\} - 1]$; (b) calculate $m_X$ using the relationship $m_Y = \exp[m_X + \sigma_Y^2/2]$; (c) derive $C_X(h)$ using the relationship $C_Y(h) = m_Y^2[\exp\{C_X(h)\}-1]$; (d) generate a homogeneous Gaussian random field $\boldsymbol{X}$ with mean $m_X$, variance $\sigma_X^2$, and covariance function $C_X(h)$; and (e) calculate $\boldsymbol{Y}$ using $\boldsymbol{Y} = \exp[\boldsymbol{X}]$.

## EXAMPLE OF RAINFALL FIELD GENERATION

By using the simulation method, predicted rainfall fields are simulated and used according to the following procedure: (a) obtain observed radar rainfall field $\boldsymbol{R}_o(t_1)$ at

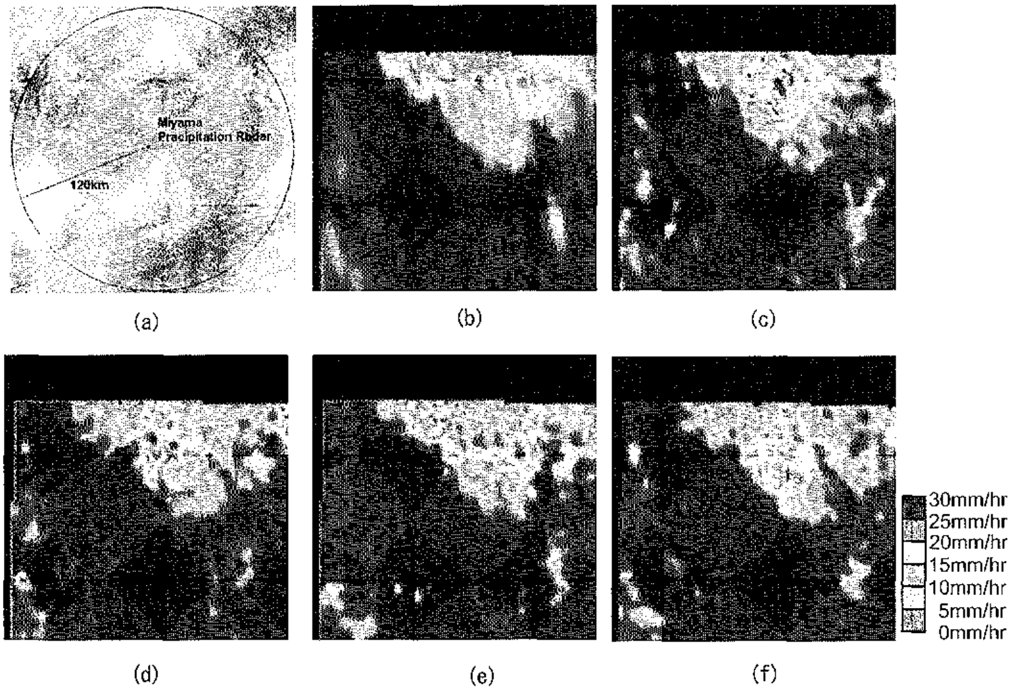


**Fig. 5** (a) Miyama radar observation area. (b) Observed 60-minute mean precipitation. (c)–(f) Generated spatial fields of 60-minute mean predicted precipitation.

current time $t_1$; (b) calculate the mean, variance and correlation length of the relative error field $\boldsymbol{E}_r(t_1)$ by using $\boldsymbol{R}_o(t_1)$ and a rainfall field $\boldsymbol{R}_p(t_1)$ predicted for $t_1$ at previous time step; (c) forecast a rainfall field $\boldsymbol{R}_p(t_2)$ for time $t_2$ ($>t_1$) by using the translation model; (d) simulate an error field $\boldsymbol{E}_r(t_2)$ with the statistical characteristics obtained at step (b); (e) calculate $\boldsymbol{R}$ by using $\boldsymbol{R}_p(t_2)$ and $\boldsymbol{E}_r(t_2)$ using the equation (4); and (f) return to step (d) and generate alternative $\boldsymbol{R}$. After generating required rainfall fields and obtaining a probability distribution of river discharge prediction though a distributed runoff model, go to next time step and return to step (a).

Figure 5 shows examples of simulated rainfall fields at 12:00 on 5 September 1989. In this case, a precipitation field with 1-h and 3-km resolution shown in Fig. 5(b) is predicted by using the translation model and realizations of rainfall fields are generated based on it. The black coloured area in the figure shows the unpredictable area because the predicted rainfall field in the area moves from the outside of the radar measured region.

## RESULTS

By using operational radar precipitation data and predicted rainfall field by the translation model, the error structure of predicted precipitation field was analysed. As a result, a relative prediction error was modelled as an isotropic lognormal spatial random field, and a method to generate a lognormal spatial random field was developed. In the method, a matrix factorization technique is adopted, where a given covariance matrix is decomposed into its square root matrix approximately by using the Chebyshev polynomials. Finally, using the method, generated realizations of rainfall fields and the simulated results were demonstrated. According to the Kolmogorov–Smirnov test of fit of a relative prediction error to a lognormal distribution, to use time or spatially-averaged radar data is suggested to apply the method presented here.

For further works, we need to devise an algorithm considering $R_p$ is zero. In the method proposed here, if a predicted rainfall in a grid cell is zero, the relative prediction error for the grid cell is supposed to be zero. Also probability distributions of predicted river discharge should be evaluated through a distributed hydrological model with the use of simulated rainfall fields.

**Acknowledgements** The authors thank Yodo River Dams Control Office, Ministry of Land, Infrastructure and Transport, Japan for providing radar rainfall data and Associate Professor Nakakita at the Department of Urban and Environmental Engineering, Kyoto University, for providing the FORTRAN source code of the translation model.

## REFERENCES

Davis, M. W. (1987) Generating large stochastic simulations—the matrix polynomial approximation method. *Math. Geol.* **19**(2), 99–107.

Georgakakos, K. P. & Krajewski, W. F. (1995) Worth of radar data in the real-time prediction of mean areal rainfall by nonadvective physically-based models. In: *New Uncertainty Concepts in Hydrology and Water Resources* (ed. by Z. W. Kundzewicz), 168–180. Cambridge University Press, Cambridge, UK.

Georgakakos, K. P. (2000) Covariance propagation and updating in the context of real-time radar data assimilation by quantitative precipitation forecast models. *J. Hydrol.* **239**, 115–129.

Shiiba, M., Takasao, T. & Nakakita, E. (1984) Investigation of short-term rainfall prediction method by a translation model. (Proc. 28th Annual Conf. on Hydraulics, JSCE) 423–428 (in Japanese). Japanese Society of Civil Engineers, Tokyo, Japan.

Takasao, T. & Shiiba, M. (1985) Development of techniques for on-line forecasting of rainfall and runoff. *Natural Dis. Sci.* **6**(2), 83–112.

Takasao, T., Shiiba, M. & Nakakita, E. (1994) A real-time estimation of the accuracy of short-term rainfall prediction using radar. In: *Stochastic and Statistical Methods in Hydrology and Environmental Engineering* (ed. by K. W. Hipel) **2**, 339–351. Kluwer Academic Publishers, The Netherlands.

Tachikwa, Y. & Shiiba, M. (2000) Gaussian and log normal random field generation based on the square root decomposition of a covariance matrix. *J. Hydraul. Coastal Environ. Engng JSCE* **656**(II-52), 39–46 (in Japanese).