Hydrol. Earth Syst. Sci. Discuss., 8, 4229–4259, 2011
www.hydrol-earth-syst-sci-discuss.net/8/4229/2011/
doi:10.5194/hessd-8-4229-2011

Hydrology and Earth System Sciences Discussions

Application of Weighted Semivariogram Model (WSVM) based on fitness to experimental semivariogram on estimation of rainfall amount

S.-J. Wu[1], P.-H. Chen[2], and J.-C. Yang[2]

[1]National Center for High-Performance Computing, Hsinchu, Taiwan
[2]Department of civil engineering, National Chiao-Tung University, Hsinchu, Taiwan

Received: 23 March 2011 – Accepted: 14 April 2011 – Published: 28 April 2011

Correspondence to: S.-J. Wu (sjwu@nchc.org.tw)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This study presents a weighted semivariogram model (WSVM) which is intended to reduce the uncertainties in the selection of the best-fit semivariogram model and associated parameters. The proposed WSVM is based on the combined forecast method, providing the weighted average semivariogram by summing up the product of estimated semivariograms and weighted factors, which are related with the inverse of the objective function value associated with the optimal parameters of theoretical semivariogram models (TSVMs). A WSVM can save computation time in the estimation of rainfall amount without the identification of the best-fit TSVM commonly carried out by the cross-validation. Ten rainstorm events recorded at fourteen rain-gauges in North Taiwan's Shinmen reservoir catchment are used to develop and validate this model by comparing the estimated rainfall amount by the Kriging method with the WSVM and TSVM, respectively. The results of the model validation indicate that the proposed WSVM not only reduces the uncertainty of failing to select the best-fit TSVM, but also effectively provides more the accurate and reliable estimated rainfall amount than TSVM.

1 Introduction

The Kriging method is a well-known geostatistical method widely applied in the estimation of hydrological variables, such as precipitation (e.g. Goovaerts, 2000; Teegavarapu and Chandramouli, 2005; Legleiter and Kyriakidis, 2008), hydraulic conductivity (e.g. Cassiani et al., 1998; Ouellon et al., 2008), soil moisture (e.g. Buttafuoco et al., 2005; Perry and Niemann, 2008), and shallow water table (e.g. Desbarats et al., 2001; Lyon et al., 2006). Moreover, the optimal number and location of rainfall gauges can be determined by the Kriging method (e.g. Pardo-Iguzauiza, 1998). The modification of the Kriging method is still in progress based on the characteristics of spatial variables. Todini (2001) presented an approximate methodology based on the truncated

HESSD
8, 4229–4259, 2011

Taylor expansion approximation in order to evaluate the influence of parameter uncertainty in the Kriging method. This model is applied in the estimation of average annual precipitation over the Veneto Region in Italy. From experiment results, this model effectively assesses the influence of parameter estimation uncertainty in the Kriging. Ortiz and Deutsch (2002) proposed an approach for the evaluation of the uncertainty in the semivariogram, and developed a methodology to transfer this uncertainty value into geostatistical simulation and decision making. Teegavarpu et al. (2005) presented a stochastic data-driven model incorporating an artificial neural network and the Kriging method for the estimation of missing precipitation records. Their results indicate that the proposed model can improve the estimation of missing precipitation and its accuracy is better than the commonly used inverse distance method. Skoien et al. (2006) suggested the topological Kriging (Top-kriging) to estimate 100-year flood in ungauged catchments in two Austrian regions. Their results proved that Top-kriging can provide more plausible and indeed more accurate estimates than Ordinary Kriging. Walter et al. (2007) developed a methodology for improving the semivariogram estimation when low sample size is applied in generating spatial autocorrelation of oyster abundance. This proposed method can reduce the likelihood of failing to obtain a variogram from a set of samples and improves the efficiency of variogram estimation.

Although several modified Kriging methods have been developed, a good best-fit theoretical semivariogram model of a spatial phenomenon is still necessary (Delay and Marsily, 1994). Delay and Marsily (1994) proposed a method of the integral of the semivariogram (ISV) to overcome the problem of grouping the pairs of experimental points into classes of distances when the data are not distributed on a regular grid. However, there are often limited data available in early stage of geostatistical modeling which leads to considerable the uncertainty in statistical parameters, including the variogram (Ortiz et al., 2002). Hence, the identification and parameter-calibration of the best-fit theoretical semivariogram model, and the estimation of spatial data, become uncertain and unreliable. The uncertainty probably results in measurement error, equipment failure, or other errors of spatial correlation and so on. Unfortunately, the

HESSD
8, 4229–4259, 2011

above uncertainty in the calculation of the variogram is hardly eliminated (Cressie and Hawkins, 1980; Genton, 1998). Although Barancourt et al. (1992) indicated that the selection of the theoretical semivariogram slightly influences the estimation of averaged monthly precipitation fields, the accuracy of predicted precipitation by the Kriging method in the high-resolution rain-gauge network is significantly affected by the theoretical semivariogram model in which the best-fit type is difficultly identified using the dispersive experimental semivariogram (Dirks et al., 1998). As with the above, Verworn and Haberlandt (2011) presented that the precipitation interpolation performance is less influenced by the effect of different semivariogram types, but its performance significantly varied with the event. In addition, the Kriging method separates two steps of the selection of the best-fit semivariogram model and the calibration of associated parameters. The selection of the best-fit semivariogram model is commonly carried out by the leave-one-out cross-validation method. In detail, cross-validation leaves one sample out and predicts for the sample location based on remaining samples (Santra et al., 2008). In doing so, the selection of the best-fit model may involve a long computation time, especially for a big sample size. Moreover, the above separation probably leads to the selection of best-fit model based on the variance of the errors of estimated data through a semivariogram model with previously calibrated parameters. Hence, the uncertainty in model parameters will influence the selection of the best-fit model (Todini, 2001).

The aim of this study is to present a weighted semivariogram model based on theoretical semivariogram models to reduce the probability of failing to select the best-fit semivariogram models and associated parameters so as to effectively produce accurate spatial estimators.

HESSD
8, 4229–4259, 2011

2 Methodology

2.1 Brief review of theoretical semivariogram model

Before introducing theoretical semivariogram models, the definition of an experimental semivariogram $\gamma(h)$ is expressed as

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x) - z(x+h)]^2 \tag{1}$$

in which h is the distance between two spatial points, $N(h)$ is the number of points within the distance h, $z(x)$ and $z(x+h)$ are spatial data at two points x and $x+h$. The commonly used theoretical semivariogram models are introduced as follows (Davis, 1973):

1. Spherical model:

$$\gamma(h) = \begin{cases} C_0 \times \left[\frac{3}{2}\frac{h}{a_0} - \frac{1}{2}\left(\frac{h}{a_0}\right)^3\right], & 0 \le h \le a_0 \\ C_0 & , h > a_0 \end{cases} \tag{2}$$

2. Exponential model:

$$\gamma(h) = C_0 \times \left[1 - \exp\left(\frac{-h}{a_0}\right)\right] \tag{3}$$

3. Gaussian model:

$$\gamma(h) = C_0 \times \left[1 - \exp\left(\frac{-h}{a_0}\right)^2\right] \tag{4}$$

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀◀ ▶▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

4. Power model:

$$\gamma(h) = C_0\, h^{a_0} \tag{5}$$

5. Linear model:

$$\gamma(h) = C_0\, h \tag{6}$$

6. Cubic model:

$$\gamma(h) = \begin{cases} C_0 \times \left[7\left(\frac{h}{a_0}\right) - \frac{35}{4}\left(\frac{h}{a_0}\right)^3 + \frac{7}{2}\left(\frac{h}{a_0}\right)^5 - \frac{3}{4}\left(\frac{h}{a_0}\right)^7\right], & 0 \le h \le a_0 \\ C_0 & , h > a_0 \end{cases} \tag{7}$$

7. Pentaspherical model:

$$\gamma(h) = \begin{cases} C_0 \times \left[\frac{15}{8}\left(\frac{h}{a_0}\right) - \frac{5}{4}\left(\frac{h}{a_0}\right)^3 + \frac{3}{8}\left(\frac{h}{a_0}\right)^5\right], & 0 \le h \le a_0 \\ C_0 & , h > a_0 \end{cases} \tag{8}$$

in which a_0 andC_0 are the influence range and the scale (or sill), respectively. This study implements the parameter calibration using the genetic algorithm method with an objective function F_{obj} as:

$$F_{\text{obj}}(m) = \left[\frac{1}{\sum_{i=1}^{n} N_{\text{p}}(h_i)} \sum_{i=1}^{n} \left(N_{\text{p}}(h_i) \times (\gamma_{\text{m}}(h_i) - \gamma_0(h_i))^2\right)\right] \tag{9}$$

in which n is the number of distance ranges and M is the number of theoretical semivariogram models. $N_{\text{p}}(h_i)$ is the number of pairs within the distance range h_i, and γ_{m} denotes the estimated semivariogram by the m-th theoretical semivariogram model, and γ_0 is the experimental semivariogram calculated using measured spatial data.

HESSD

8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.

2.2 Development of a weighted semivariogram model

2.2.1 Model concept

To reduce the uncertainty of failing to select the best-fit theoretical semivariogram model and associated parameters, this study refers to the combined forecasts method (Fischer and Harvey, 1999) to develop a weighted semivariogram model. The combined forecasts method is a well-established procedure to improve forecasting accuracy which takes advantage of the availability of both multiple information and computing resources for data-intensive forecasting (Bunn, 1989). Basically, the proposed weighted semivariogram model combines results from theoretical semivariogram models to provide the weighted average of semivariogram $\gamma_{\mathrm{w}}(h)$ by using the following equation:

$$\gamma_{\mathrm{w}}(h) = \sum_{m=1}^{M} w_{\mathrm{sv}}(m) \times \gamma_{\mathrm{m}}(h) \tag{10}$$

in which w_{sv} is the weighted factor and $\gamma_{\mathrm{m}}(h)$ denotes the estimated semivariogram estimated m-tj theoretical semivariogram model. In view of Eq. (7), the objective function value F_{obj} decreases with the error $\left[\gamma_{\mathrm{m},i}(h) - \gamma_{\mathrm{o},i}(h)\right]^2$. Hence, the less objective function value indicates that the estimated semivariogram $\gamma_{\mathrm{m}}(h)$ fits better to the experimental one $\gamma_{\mathrm{o}}(h)$, that is, the fitness of estimated runoff to experimental data increases. Accordingly, if a theoretical semivariogram model has the minimum objective function F_{obj} among theoretical models, theoretically, it can produce a more an accurate semivariogram. As a result, the accuracy of estimated semivariogram by theoretical models is inversely proportion to F_{obj}, namely, it is positively related with the inverse of F_{obj}. Therefore, this study defines the weighted factor w_{sm} as a function of the inverse of the objective function value F_{obj} and can be calculated by the following equation:

HESSD
8, 4229–4259, 2011

$$w_{sv}(m) = \frac{\frac{1}{F_{obj}(m)}}{\sum_{m=1}^{M}\left\{\frac{1}{F_{obj}(m)}\right\}} \tag{11}$$

where $F_{obj}(m)$ denotes the objective function for the m^{th} theoretical semivariogram model associated with the optimal model parameters. Note that the sum of w_{sv} should be equal to one. Figure 1 shows the graphical illustration of theoretical semivariogram models and the weighted semivariogram model.

Substituting the weighted semivariogram $\gamma_w(h)$ into the Kriging equation system, the Kriging weight λ can be solved

$$\begin{aligned} \gamma_w\left(h_{0,i}\right) &= \sum_{j=1}^{K}\left[\lambda_j \times \gamma_w\left(h_{i,j}\right)\right] + \mu \\ \sum_{j=1}^{K}\lambda_j &= 1 \end{aligned} \tag{12}$$

where μ is the Lagrange multiplier. $h_{0,j}$ is the distance between the point x_0 of which data $\hat{z}(x_0)$ would be estimated as well as point x_j; of which data $z(x_j)$ are known, and K is the number of locations measured. Eventually, the estimate at the point $x_0\hat{z}(x_0)$ can be arrived at d through the following equation with the measured data at points $x_j z(x_j)$.

$$\hat{z}(x_0) = \sum_{j=1}^{K}\left(\lambda_j \times z\left(x_j\right)\right) \tag{13}$$

2.2.2 Development procedure

To derive the proposed weighted semivariogram model, the procedure of model development is expressed as:

HESSD
8, 4229–4259, 2011

- Step 1: Calculate the experimental semivariogram.
- Step 2: Calibrate the parameters of theoretical semivariogram models using the experimental semivariogram.
- Step 3: Calculate the weighted factors of theoretical semivariogram models using Eq. (11).
- Step 4: Estimate the semivariograms using the theoretical semivariogram models and calculate the corresponding weighted average through Eq. (10)
- Step 5: Solve the Kriging equation system composed of the weighed theoretical semivariogram to obtain the Kriging weights using Eq. (12) and then predict the spatial data at the objective points through Eq. (13).

2.3 Model validation

In this study, the model validation is made by comparing the estimated rainfall amount at rain-gauges by the Kriging method with weighed and theoretical semivariogram models. To investigate the effect of sample size on the estimation of rainfall amount, the cross validation method is implemented in the mode validation. In detail, some rain-gauges are randomly extracted from the catchment area, which are defined as calibration gauges, and the remaining gauges are regarded as validation gauges used for the model development and validation. Then, the rainfall amounts at validation gauges are estimated using TVMS and WSVM, in which the associated parameters are calibrated using the observed data at calibration gauges, and the associated model performance indices are calculated for the model assessment. Note that the weighted and theoretical semivariogram models are named WSVM and TSVM respectively in this study.

HESSD
8, 4229–4259, 2011

The performance indices commonly used in the model validation are introduced as:

1. Root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{val}}}\sum_{i=1}^{N_{\mathrm{val}}}\left(R_{\mathrm{est},i} - R_{\mathrm{obs},i}\right)^2} \qquad (14)$$

   in which N_{val} is the number of validation gauges, R_{est} and R_{obs} are estimated and observed rainfall amount at the validation gauges estimated using Kriging method with TSVMs or WSVM, respectively. Note that a small RMSE value indicates that the estimated rainfall amount could be closer to the observed one.

2. Model reliability index (KG) (Leggett and Williams, 1981):

$$KG = \frac{1 + \sqrt{\frac{1}{N_{\mathrm{val}}}\sum_{i=1}^{N_{\mathrm{val}}}\left[\frac{1-\left(\frac{R_{\mathrm{est},i}}{R_{\mathrm{obs},i}}\right)}{1+\left(\frac{R_{\mathrm{est},i}}{R_{\mathrm{obs},i}}\right)}\right]^2}}{1 - \sqrt{\frac{1}{N_{\mathrm{val}}}\sum_{i=1}^{N_{\mathrm{val}}}\left[\frac{1-\left(\frac{R_{\mathrm{est},i}}{R_{\mathrm{obs}^i}}\right)}{1+\left(\frac{R_{\mathrm{est},i}}{R_{\mathrm{obs},i}}\right)}\right]^2}} \qquad (15)$$

   Note that KG approaching one implies that the spatial variation of the estimated rainfall amount could resemble observed one.

3. Probability of performance indices of estimated rainfall amount using WSVM less than those for TVSM.

Since calibration gauges are randomly selected by means of the bootstrap method in this study, the resulting performance index RMSE and KG values are probably dependent on the locations and number of calibration gauges extracted. In addition to comparison of RMSE and *KG*, this study evaluates WSVM and TSVM by calculating probability of performance indices for WSVM superior to those for TSVM. Thus, using a number of estimated rainfall amounts by means of WSVM and TSVM, the corresponding probability $\Pr(\text{RMSE}_{\text{WSVM}} < \text{RMSE}_{\text{TSVM}})$ is calculated. Similarly, the probability of WSVM closer to one than that for TSVMs $\Pr((KG_{\text{WSVM}} - 1) < (KG_{\text{TSVM}} - 1))$ is computed. By comparing $\Pr(\text{RMSE}_{\text{WSVM}} < \text{RMSE}_{\text{TSVM}})$ and $\Pr((KG_{\text{WSVM}} - 1) < (KG_{\text{TSVM}} - 1))$ the effect of the number and location of calibration gauges to WSVM and TSVM in the estimation of rainfall amounts at ungauged zones can be analyzed, and the results can be referred to the evaluation of the proposed WSVM.

3 Results and discussion

In this section, the estimated rainfall amounts are produced by the Kriging method with the best-fit TSVM and WSVM and compared using the performance indices calculated. Since the estimation of rainfall amount probably is probably influenced by rainstorm events and number, as well as location, of rain-gauges, the model validation has two parts: one is to consider the uncertainty of rainstorm events and the other is to take into account the effects of number and locations of rain-gauges on the estimation of rainfall amount.

3.1 Study area and data used

For the model development and validation, the Shinmen reservoir catchment is adopted as the study area (see Fig. 2). The Shinmen Reservoir is located upstream of the Dahan River basin in northern Taiwan, and serves a number of purposes, including irrigation, hydroelectric power, fresh water supply, flood prevention and sightseeing. Ten rainstorm events recorded at the fourteen rain-gauges from 2004 to 2008 in Shinmen reservoir catchment are used as the study data, as shown in Tables 1 and 2.

HESSD
8, 4229–4259, 2011

3.2 Consideration of uncertainty in rainstorm events

To reflect the uncertainty in rainstorm events, the observed rainfall amounts of ten rainstorm events (see Table 2) are used in the model validation. Note that seven gauges of the fourteen rain-gauges in the Shinmen reservoir catchment are selected as calibration gauges, whereas the remaining gauges are validation gauges (see Table 1).

3.2.1 Identification of best-fit TSVM

Using the Kriging method to estimate spatial data, the best-fit TSVM should be identified in advance. In general, the leave-one-out cross-validation method is widely applied in the identification of the best-fit model based on the standardized average error (SKAE) and the standardized kriging variance (SKV) (Evrendilek and Frtekin, 2007; Kumar and Remadevi, 2006) as:

$$\mathrm{SKAE} = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{\left(z^{*}(x_i) - z(x_i)\right)}{\sigma_{\mathrm{k}i}}\right] \tag{16}$$

$$\mathrm{SKV} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[\frac{\left(z^{*}(x_i) - z(x_i)\right)}{\sigma_{\mathrm{k}i}}\right]^2} \tag{17}$$

where N is the sample size. $z^{*}(x_i)$ and $z(x_i)$ are the estimated and observed data at the point x_i. $\sigma_{\mathrm{k}i}$ denotes the estimated kriging variance at the point x_i. SKAE accounts for an indicator of prediction errors, which means the degree of bias in model prediction, and it is required to be close to zero. Moreover, SKAE is supposed to be in the range $1 \pm 2\sqrt{2N}$. SKV reveals the comparison of the error variance to the kriging variance, and should be close to one. SKV greater and less than one means that the predictions are underestimated and overestimated, respectively. In summary, the best-fit TSVM should satisfy the criteria, i.e. $\mathrm{SKAE} \cong 0$ and $\mathrm{SKV} \cong 1$. In addition, before calculating

HESSD
8, 4229–4259, 2011

the experimental semivariogram used in the identification of the best-fit model and parameter-calibration, the rainfall amount should be non-dimensionalized through the equation $R^*(x) = R(x)/\sigma_R$in which $R^*(x)$ is a dimensionless value of rainfall amount $R(x)$ and σ_R is the standard deviation of rainfall amounts.

Table 3 lists SKAE and SKV values of TVSMs for eight rainstorm events. As shown, TSVM has significantly different SKAE and SKV values for rainstorm events, so the selected best-fit TSVM varies with the rainstorm event, based on the abovementioned criteria. Specifically, Power and Gaussian models have the maximum SKAE (on average 4.09) and SKV (on average 18.86) respectively, which indicates that Power and Gaussian models are unlikely to be selected as the best-fit models. According to SKAE, the best-fit TSVMs for rainstorm events are Spherical (EV2 and EV4), Exponential (EV1), Gaussian (EV5, EV6, EV10), Linear (EV3), Cubic (EV7), and Pentaspherical (EV8). However, referring to SKV, the best-fit TSVMs are Spherical (EV1 and EV5), Exponential (EV2, EV9, and EV10), Power (EV3 and EV8), Linear (EV7), and Cubic (EV6). In summary, the Spherical, Exponential, and Gaussian models are frequently identified as the best-fit models.

Although the best-fit TSVM can be determined based on SKAE and SKV as shown in Table 1, it is observed that the corresponding SKAE and SKV values are obviously greater than zero and one, respectively. This implies the best-fit model has a significant model bias and the resulting predictions may be underestimated.

3.2.2 Calculation of weighted factors of TSVM

As with deriving WSVM, the weight factors of TSVMs should be calculated in advance. According to Eq. (10), the weighted factors are based on the objective function values associated with the optimal parameters of TSVMs. Using the genetic algorithm with the objection function Eq. (8) in this study, the optimal parameters of TSVMs can be calibrated, and the corresponding objective function values are computed as shown in Table 4. Thus, the weighted factors of TSVM used in WSVM can be quantified (see Table 5) and then the weighted semivariogram can be calculated. On average, the

HESSD
8, 4229–4259, 2011

average weighted factors are 0.152 (Spherical), 0.144 (Exponential), 0.147 (Gaussian), 0.097 (Power), 0.136 (Linear), 0.177 (Cubic), and 0.147 (Pentaspherical).

Therefore, except for the Power model, the weighted factors of TSVMs approximate 0.15. It can be also said that TSVMs (except for the Power model) may provide equivalent contributions to the estimated semivariogram so that the estimated semi-variogram by all TSVMs should be taken into account. This result differs from the consequences for SKAE and SKV criteria in which Spherical, Exponential, and Gaussian models are identified as the best-fit models. As a result, theoretically, the proposed WSVM could reduce the uncertainty in the selection of the best-fit TSVM so as to enhance the reliability of estimated rainfall amount.

3.2.3 Comparison of estimated rainfall amount

Through the Kriging system equation associated with estimated semi-variogram by the best-fit models, which are determined based on SKAE and SKV criteria, and WSVM for ten rainstorm events, the rainfall amounts at seven validation points are estimated. Figure 3 shows the graphical comparison of observed and the estimated rainfall amounts at validation gauges by the best-fit TSVMs and WSVM. It can be observed that the estimated rainfall amount by WSVM significantly differs from those by the best-fit TSVMs for ten rainstorm events. Although the fitness of estimated rainfall amount to observed data varies with the rainstorm event, the estimated rainfall amount visually fits the observed data better than those by the best-fit TSVMs, except for EV2 and EV7 in which the estimated rainfall by WSVM resembles those by the best-fit TSVMs. Therefore, it is shown that WSVM could capture the behavior of rainfall amount better than the best-fit TSVMs.

The performance index RMSE values are also calculated with the observed and estimated rainfall amounts at validation gauges by WSVM and the best-fit TSVM as shown in Fig. 4. From Fig. 4, the RMSE values of estimated rainfall amounts by WSVM are less than or approximately equal to those by the best-fit TSVM. Specifically, the average RMSE for WSVM (1.372) is significantly less than those for the best-fit TSVMs,

HESSD
8, 4229–4259, 2011

i.e. 1.458 (SKAE) and 1.508 (SKV). It follows that the proposed WSVM could effectively provide more accurate estimation of rainfall amount than the best-fit models identified by SKAE and SKV criteria by taking into account the estimated semivariogram by TSVMs.

In addition to the performance index RMSE, this study also calculates the model reliability index *KG* to evaluate the spatial variation of estimated rainfall amount (see Fig. 5). From Fig. 5, similar to RMSE, the *KG* values of estimated rainfall amount by WSVM are less or approximately equal to those by the best-fit TSVM. On average, the *KG* value of estimated rainfall amount by WSVM approximates 1.372 which is less than those by the best-fit TSVMs, i.e. 1.458 (SKAE) and 1.508 (SKV). It can be also said that the *KG* values of estimated rainfall amounts by WSVM are closer to one than those by the best-fit TSVMs. Therefore, WSVM could capture the behavior of rainfall amount better than the best-fit TSVM.

To sum up the above results, the proposed WSVM can reduce the uncertainty resulting from unsuitable best-fit TSVMs so as to produce more accurate and reliable estimated rainfall amount.

3.3 Consideration of uncertainty in number and location of rain-gauges

To evaluate the effect of number and locations of rain-gauges on the estimation of rainfall amount by the Kriging system equation with WSVM and the best-fit TSVM, 4–11 rain-gauges are randomly extracted as calibration gauges and the remaining gauges are defined as the validation gauges by fifty times. According to results from Table 1, Spherical, Gaussian, and Exponential models have a high likelihood of being selected as the best-fit model. Therefore, this study adopts Spherical, Gaussian, and Exponential models as the best-fit TSVMs and compares them with WSVM in the estimation of rainfall amount under the consideration of various number and locations of calibration gauges.

Figure 6 shows the comparison of average RMSE of estimated rainfall amount by WSVM and the best-fit models with various number of calibration points. The average

HESSD
8, 4229–4259, 2011

RMSE values of estimated amount by WSVM are significantly less than those by the best-fit TSVMs with various number of calibration points. Overall, the average RMSE value for WSVM (about 140.1) is less than those for the best-fit models, i.e. 150.4 (Spherical), 144.1 (Exponential), and 150.8 (Gaussian). Similar to RMSE, the average *KG* values of estimated rainfall amount by WSVM (averagely 1.485) are closer to one than those by the best-fit TSVMs, i.e. 1.535 (Spherical), 1.511 (Exponential), and 1.533 (Gaussian). Therefore, in considering the uncertainty in the number and location of rain-gauges, WSVM can produce the more accurate and reliable rainfall amounts than TSVMs.

Referring to the above results, it is shown that WSVM can provide the rainfall amount associated with less RMSE and *KG* values than TSVMs. However, the above conclusion is based only on the tendency of average RMSE and *KG* values. To completely investigate fifty sets of RMSE and *KG* values, the probabilities $\Pr(\mathrm{RMSE}_{\mathrm{WSVM}} < \mathrm{RMSE}_{\mathrm{TSVM}})$ and $\Pr\left[(KG_{\mathrm{WSVM}}-1) < (KG_{\mathrm{TSVM}}-1)\right]$ are calculated as shown in Fig. 7. In view of Fig. 7, $\Pr(\mathrm{RMSE}_{\mathrm{WSVM}} < \mathrm{RMSE}_{\mathrm{TSVM}})$ for different number of calibration gauges are mostly greater than 50 %. Specifically, the average values of $\Pr(\mathrm{RMSE}_{\mathrm{WSVM}} < \mathrm{RMSE}_{\mathrm{TSVM}})$ are 63.4 % (Spherical), 61.6 % (Gaussian), and 55.6 % (Exponential), respectively. As for $\Pr\left[(KG_{\mathrm{WSVM}}-1) < (KG_{\mathrm{TSVM}}-1)\right]$, t he average for Spherical, Gaussian, and Exponential models are 62.4 %, 62.9 %, and 51.7 % , respectively. In the cases of Spherical and Gaussian models being the best-fit models, WSVM can produce the estimation of rainfall amounts with a corresponding probability of 60 % which are more accurate and reliable than the best-fit models. Even for the Exponential model, which has fewer probabilities $\Pr(\mathrm{RMSE}_{\mathrm{WSVM}} < \mathrm{RMSE}_{\mathrm{TSVM}})$ and $\Pr\left[(KG_{\mathrm{WSVM}}-1) < (KG_{\mathrm{TSVM}}-1)\right]$ as compared to the Spherical and Gaussian models, WSVM has a 50 % probability of capturing the behavior of estimated rainfall amount in scale and space better than the Exponential model.

In summary, varying according to number and location of rain-gauges, WSVM has a high likelihood of capturing the rainfall amount at validation gauges, that is, WSVM is significantly superior to the best-fit model in the estimation of rainfall amount under

HESSD
8, 4229–4259, 2011

the consideration of uncertainty in the number and location of rain-gauges. In addition, since WSVM estimates the rainfall amount without identifying the best-fit TSVM through cross-validation, it takes less computation time and is more effective than TSVM.

4 Conclusions

This study proposes a weighted semivariogram model (WSVM) to reduce the uncertainty in the identification of the best-fit theoretical semivariogram models (TSVMs) and associated parameters. The proposed WSVM mainly calculates the weighted average of the semivariogram, which is the sum of product of estimated semivariogram and weighted factors resulting from the inverse of the objective function value associated with optimal parameters. WSVM can effectively produce the rainfall amount without determining the best-fit TSVM commonly carried out by the cross-validation method. The results of the graphical comparison and performance indices for ten rainstorm events recorded in the Shinmen reservoir catchment indicate the proposed WSVM not only improves the accuracy of estimated rainfall amount without determining the best-fit model, but also has a high probability of capturing the real rainfall amount under the consideration of uncertainties in rainstorm events and number as well as locations of rain-gauges. Consequently, it is shown that the WSVM can effectively reduce the uncertainty of failing to select an unsuitable model and provide the more accurate and reliable rainfall amount.

A number of future investigations will be performed, in which WSVM can be incorporated with the other Kriging methods, such as the Indicator Kriging (IK), Universal Kriging (UK), Disjunctive Kriging (DK) and Top-Kriging (TK) and so on, to be applied in other study areas of interest. In addition, WSVM will be extended to become a weighted spatio-temporal semivariogram model and applied in the estimation of spatio-temporal data, such as the rainfall hyetograph or rainstorm patterns.

HESSD
8, 4229–4259, 2011

References

Barancourt, C., Creutin, J. D., and Rivoirard, J.: A Method for Delineating and Estimating Rainfall Fields, Water Resour. Res., 28(4), 1133–1144, 1992.

Buttafuoco, G. and Castrignano, A.: Study of the Spatio-Temporal Variation of Soil Moisture under Forest Using Intrinsic Random Functions of Order *k*, Geoderma, 128(3), 208-220, 2005.

Bunn, D. W.: (Forecasting More than One model, J. Forecast., 8, 161–166, 1998.

Cassiani, G., Böhmd, G., Vesnaverd, A., and Nicolichd, R.: A Geostatistical Framework for Incorporating Seismic Tomography Auxiliary Data into Hydraulic Conductivity Estimation, J. Hydrol., 206, 58–74, 1998.

Cressie, N. and Hawkins, D.: Robust Estimation of the Variogram, Math. Geol., 12(2), 115–126, 1980.

Davis, J. C.: Statistics and data analysis in geology, John Wiley & Sons, New York, 1973.

Delay, F. and de Marsily, G.: The Integral of the Semivariogram: A Powerful Method for Adjusting the Semivariogram in Geostatistics, Math. Geol., 26(3), 301–321, 1994.

Desbarats, A. J., Logan, C. E., Hinton, M. J., and Sharpe, D. R.: On the Kriging of Water Table Elevations Using Collateral Information from A Digital Elevation Model, J. Hydrol., 255, 25–38, 2002.

Evrendilek, F. and Ertekin, C.: Statistical modeling of spatio-temporal variability in monthly average daily solar radiation over Turkey, Sensors, 7, 2763–2778, 2007.

Fischer, I. and Harvey, H.: Combing forecasts: What information do judges need to outperform the simple average?, Int. J. Forecast., 15, 227–246, 1999.

Genton, M. G.: Highly Robust Variogram Estimation, Math. Geol., 30(2), 213–221, 1998.

Goovaerts, P.: Geostatistical Approaches for Incorporating Elevation into the Spatial Interpolation of Rainfall, J. Hydrol., 228, 113–129, 2000.

Kumar, V. and Remadevi, V.: Kriging of groundwater level – a case study, J. Spatial Hydrol., 6(1), 81–92, 2006.

Leggett, R. W. and Williams, L. R.: A reliability index for models, Ecol. Modell., 13, 303–312, 1981.

Legleiter, C. J. and Kyriakidis, P. C., , Spatial Prediction of River Channel Topography by Kriging, Earth Surf. Proc. Land., 33(6), 841–867, 2008.

Lyon, S. W., Seibert, J., Lembo, A. J., Walter, M. T., and Steenhuis, T. S.: Geostatistical investigation into the temporal evolution of spatial structure in a shallow water table, Hydrol. Earth Syst. Sci., 10, 113–125, doi:10.5194/hess-10-113-2006, 2006
Ortiz, J. M. and Deutsch, C.: Calculation of Uncertainty in the Variogram, Math. Geol., 34(2), 169–183, 2002.
Ouellon, T., Lefebvre, R., Marcotte, D., Boutin, A., Blais, V., and Parent, M.: Hydraulic Conductivity Heterogeneity of a Local Deltaic Aquifer System From the Kriged 3D Distribution of Hydrofacies from Borehole Logs, Valcartier, Canada, J. Hydrol., 351(1), 71–86, 2008.
Pardo-Iguzquiza, E.: Optimal Selection of Number and Location of Rainfall Gauges for Areal Rainfall Estimation Using Geostatistics and Simulated Annealing, J. Hydrol., 210, 206–220, 1998.
Perry, M. A. and Niemann, J. D.: Generation of soil moisture patterns at the catchment scale by EOF interpolation, Hydrol. Earth Syst. Sci., 12, 39–53, doi:10.5194/hess-12-39-2008, 2008.
Santra, P., Chopra, U. K., and Chskraborty, D.: Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm, Current Sci., 95(7), 937–945, 2008.
Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, Hydrol. Earth Syst. Sci., 10, 277–287, doi:10.5194/hess-10-277-2006, 2006.
Teegavarapu, R. S. V. and Chandramouli, V.: Improved Weighting Methods, Deterministic and Stochastic Data-Driven Models for Estimation of Missing Precipitation Records, J. Hydrol., 312(1), 191–206, 2005.
Todini, E.: Influence of parameter estimation uncertainty in Kriging: Part 1 – Theoretical Development, Hydrol. Earth Syst. Sci., 5, 215–223, doi:10.5194/hess-5-215-2001, 2001.
Todini, E., Pellegrini, F., and Mazzetti, C.: Influence of parameter estimation uncertainty in Kriging: Part 2 – Test and case study applications, Hydrol. Earth Syst. Sci., 5, 225–232, doi:10.5194/hess-5-225-2001, 2001.
Verworn, A. and Haberlandt, U.: Spatial interpolation of hourly rainfall - effect of additional information, variogram inference and storm properties, Hydrol. Earth Syst. Sci., 15, 569–584, doi:10.5194/hess-15-569-2011, 2011.
Walter III, J. F., Hoenig, J. M., and Mann, R.: Combining Data from Multiple Years or Areas to Improve Variogram Estimation, Environmetrics, 18(6), 583–598, 2007.

HESSD
8, 4229–4259, 2011

Table 1. Information on the rain-gauges in study area Shinmen reservoir catchment.

Gauge	Name	Location		Type
		X	Y	
R1	Shinmen	273 867.9	2 745 779	Calibration point
R2	zhang-xing	280 331	2 743 668	Validation Point
R3	Fu-xing	285 381.5	2 745 526	Calibration Point
R4	Xia-yun	285 386.1	2 743 680	Calibration Point
R5	Gao-yi	285 409.9	2 734 450	Validation Point
R6	San-guang	287 110.9	2 728 917	Calibration Point
R7	Ga-la-he	290 495.4	2 725 234	Validation Point
R8	Yu-feng	280 363.4	2 728 900	Validation Point
R9	Xiu-luan	278 703.1	2 715 975	Validation Point
R10	Zhen-xi-bao	280 387.5	2 717 825	Calibration Point
R11	Ba-ling	288 792.6	2 730 767	Validation Point
R12	Bai-shi	271 949.4	2 715 963	Calibration Point
R13	Xi-yue-si-shan	285 447.5	2 719 683	Validation Point
R14	Chi-duan	297 238.4	2 727 102	Calibration Point

HESSD
8, 4229–4259, 2011

HESSD
8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.

Table 2. Observed rainfall amount of ten rainstorm events recorded at fourteen rain-gauges in Shinmen reservoir catchment.

Event Rainfall(mm) Gauge	EV1 20 040 823 ~ 200 400 826	EV2 20 050 716 ~ 20 050 720	EV3 20 050 803 ~ 20 050 806	EV4 20 050 830 ~ 20 050 901	EV5 20 050 909 ~ 20 050 911	EV6 20 050 921 ~ 20 050 923	EV7 20 050 930 ~ 20 051 003	EV8 20 080 726 ~ 20 080 729	EV9 20 080 911 ~ 20 080 916	EV10 20 080 926 ~ 20 080 629
R1	437	622	127	286	56	39	199	459	316	531
R2	410.5	774	167.5	134	111.5	91.5	206.5	651.5	512	595
R3	507.5	889.5	192.5	185.5	154.5	107	250	711	577	595
R4	443	892	271	207	206	114	249	766	502	923
R5	460	1085	239	192	63	134	427	1003	454	1214
R6	343.5	888	215	230.5	71	110.5	417	930.5	516	1217.5
R7	389	861	239	224	105	80	403	676	429	857
R8	266	1096	135	172	28	152	328	1274	405	1578
R9	312.5	908	208.5	185.5	21.5	39	339.5	1109.5	443	1316.5
R10	349	926	243	197	44	98	395	872	451	1073
R11	476	1079	277	262	117	107	529	685	688	789
R12	311	1000	200	172	26	99	341	1263	437	1607
R13	617	980	555	348	102	59	456	529	707	655
R14	146	282	431	257	194	75	508	703	610	923

Table 3. SKAE and SKV of TSVMs in the cross-validation for determination of best-fit model.

(1) Standardized kriging average error SKVE							
Event	Spherical	Exponential	Gaussian	Power	Linear	Cubic	Pentaspherical
EV1	4.569	3.636*	4.218	8.473	4.630	4.671	4.134
EV2	3.130*	3.744	3.522	5.581	3.544	3.235	3.558
EV3	2.625	2.080	2.137	2.735	1.938*	2.346	2.677
EV4	4.746*	5.641	5.156	4.821	6.139	5.920	5.587
EV5	1.311	1.197	0.747*	2.718	0.979	1.179	1.731
EV6	3.638	2.518	−0.256*	1.103	4.579	2.739	5.892
EV7	3.650	3.447	3.586	4.909	3.375	3.068*	3.603
EV8	3.088	3.308	3.223	3.537	4.196	7.284	2.214*
EV9	5.316	5.568	5.839	3.961*	5.560	5.254	5.110
EV10	3.056	2.829	2.453*	3.079	-2.805	2.839	2.651
(2) Standardized kriging variance SKV							
Event	Spherical	Exponential	Gaussian	Power	Linear	Cubic	Pentaspherical
EV1	4.859*	6.462	5.780	44.840	5.543	7.469	7.158
EV2	7.392	4.417*	6.262	16.293	4.029	6.429	7.552
EV3	2.928	4.510	3.705	2.740*	4.087	3.129	2.769
EV4	10.842	5.994	9.851	30.301	6.331	6.704	6.350
EV5	1.495*	1.575	2.302	8.730	2.501	1.574	2.111
EV6	16.662	4.607	72.371	22.673	8.416	3.706*	47.275
EV7	4.469	4.283	4.710	28.821	3.947*	4.030	3.761
EV8	6.186	4.823	5.432	4.563*	5.784	106.397	14.238
EV9	6.364	5.659*	8.597	25.041	5.728	6.709	8.309
EV10	3.548	3.925*	5.170	4.548	375.773	4.253	7.200
EV9	6.364	5.659*	8.597	25.041	5.728	6.709	8.309
EV10	3.548	3.925*	5.170	4.548	375.773	4.253	7.200
Mean	6.475	4.626	12.418	18.855	42.214	15.040	10.672

Note: *Stands for the best-fit model.

HESSD
8, 4229–4259, 2011

Table 4. Optimal parameters of theoretical semivariogram models and associated objective function F_{obj}.

Event	Parameter	Spherical	Exponential	Gaussian	Power	Linear	Cubic	Pentaspherical
EV1	a_0	4704.7	7724.9	5090.4	0.3	17 028.1	6878.6	7194.3
	C_0	1.196	1.706	1.173	0.207		1.328	1.181
	F_{obj}	0.081	0.088	0.093	0.102	0.103	0.080	0.091
EV2	a_0	15062.2	7662.2	9229.2	0.3	34 038.9	11 588.9	9476.6
	C_0	1.675	1.869	1.751	0.238		1.469	1.390
	F_{obj}	0.121	0.126	0.127	0.123	0.141	0.076	0.124
EV3	a_0	16 228.0	11 031.8	11 586.2	0.2	36 771.2	12 187.0	8058.2
	C_0	2.106	2.057	2.147	0.324		1.470	1.371
	F_{obj}	0.053	0.069	0.068	0.091	0.082	0.040	0.063
EV4	a_0	28 410.0	12 877.3	8993.0	0.2	13 189.0	23 122.1	9889.9
	C_0	1.885	1.913	1.607	0.200		1.617	1.231
	F_{obj}	0.101	0.102	0.100	0.131	0.102	0.100	0.101
EV5	a_0	17 805.6	11 536.4	9757.7	0.2	18 567.9	16 338.3	8161.6
	C_0	1.605	1.924	1.954	0.335		1.432	1.307
	F_{obj}	0.058	0.060	0.053	0.098	0.057	0.054	0.057
EV6	a_0	5205.9	5986.5	5093.4	0.2	29 222.1	6947.7	15 109.3
	C_0	0.954	1.509	0.858	0.282		0.950	1.667
	F_{obj}	0.114	0.120	0.123	0.125	0.124	0.114	0.122
EV7	a_0	14 771.7	14 883.9	6855.1	0.2	16 383.9	15 258.8	9633.1
	C_0	2.236	2.880	1.450	0.229		1.451	1.274
	F_{obj}	0.042	0.040	0.037	0.071	0.050	0.024	0.040
EV8	a_0	15 559.8	15 358.6	5057.7	0.2	23 924.7	16 259.5	25 698.9
	C_0	0.608	1.046	0.573	0.415		0.831	1.504
	F_{obj}	0.045	0.045	0.045	0.138	0.045	0.045	0.045
EV9	a_0	18913.9	13 440.5	9195.4	0.2	43 758.9	22 643.1	6881.6
	C_0	1.496	1.715	1.450	0.353		1.564	1.198
	F_{obj}	0.114	0.115	0.113	0.137	0.114	0.113	0.114
EV10	a_0	6242.0	7074.5	3809.1	0.3	27 864.1	5739.9	11 342.7
	C_0	0.346	0.422	0.250	0.374		0.274	1.336
	F_{obj}	0.033	0.034	0.034	0.240	0.035	0.032	0.033

Note: a_0 and C_0 are the influence range (m) and sill (mm^2).

HESSD

8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.

Table 5. Weighted factors of theoretical semivariogram models (TSVMs) for proposed weighted semivariogram model (WSVM).

Event	Spherical	Exponential	Gaussian	Power	Linear	Cubic	Pentaspherical
EV1	0.160	0.146	0.139	0.127	0.125	0.162	0.142
EV2	0.137	0.131	0.129	0.134	0.117	0.218	0.134
EV3	0.170	0.129	0.132	0.098	0.109	0.222	0.141
EV4	0.147	0.146	0.149	0.114	0.146	0.149	0.148
EV5	0.149	0.143	0.161	0.088	0.150	0.159	0.151
EV6	0.150	0.143	0.139	0.137	0.138	0.151	0.141
EV7	0.134	0.142	0.154	0.080	0.113	0.235	0.142
EV8	0.158	0.158	0.158	0.051	0.158	0.158	0.158
EV9	0.146	0.145	0.147	0.122	0.147	0.148	0.146
EV10	0.165	0.161	0.161	0.023	0.156	0.168	0.166

HESSD
8, 4229–4259, 2011

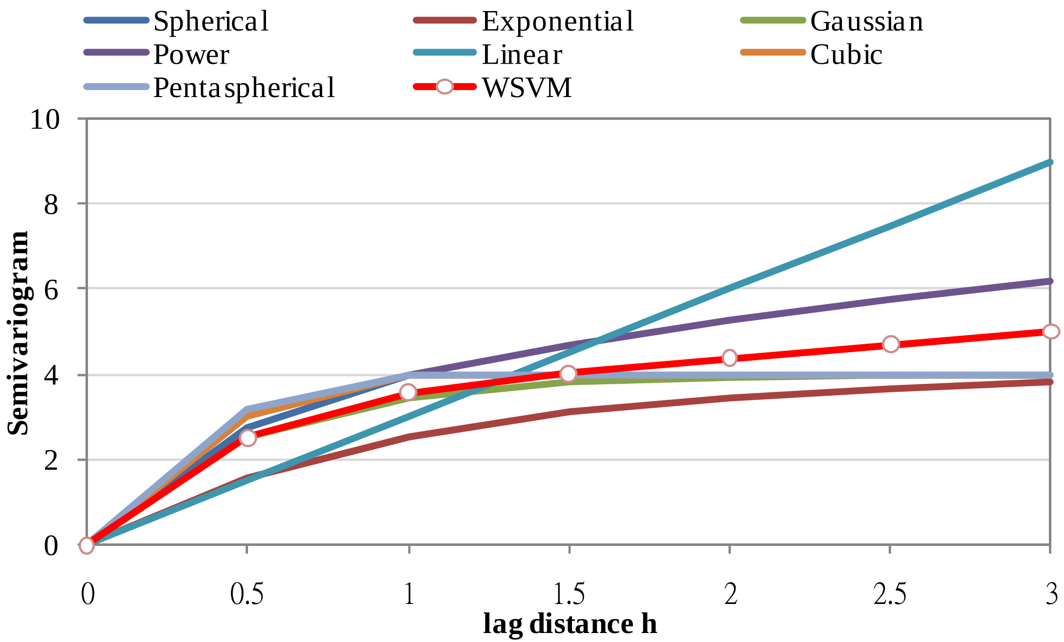

Fig. 1. Graphical illustration for semivariogram models (TSVMs) and weighed semivariogram model (WSVM).

HESSD

8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◂ ▸
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

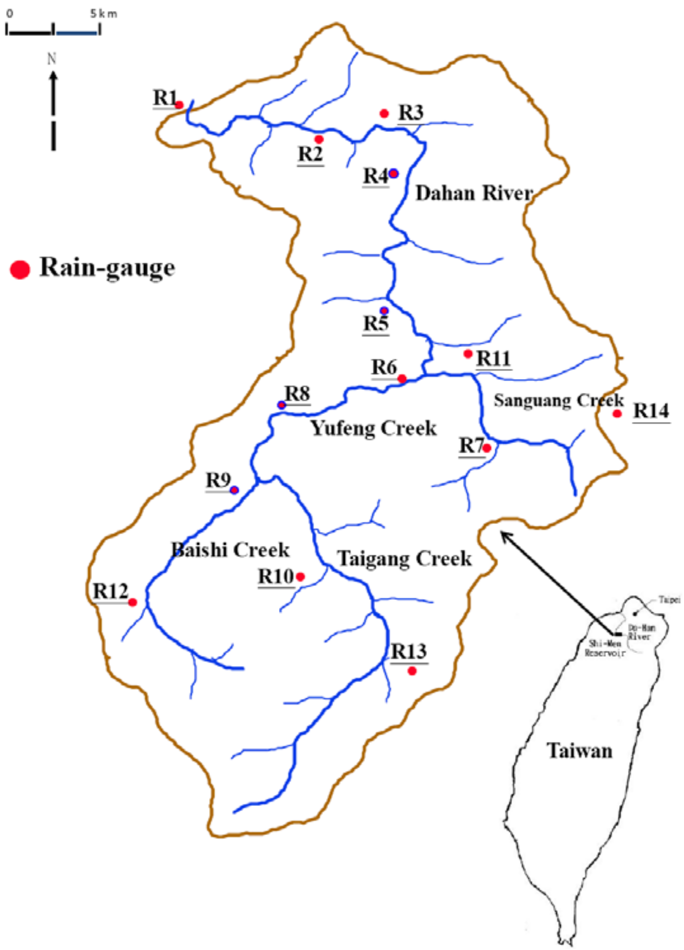

Fig. 2. Locations of fourteen rain-gauges in Shinmen reservoir catchment.

HESSD
8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.

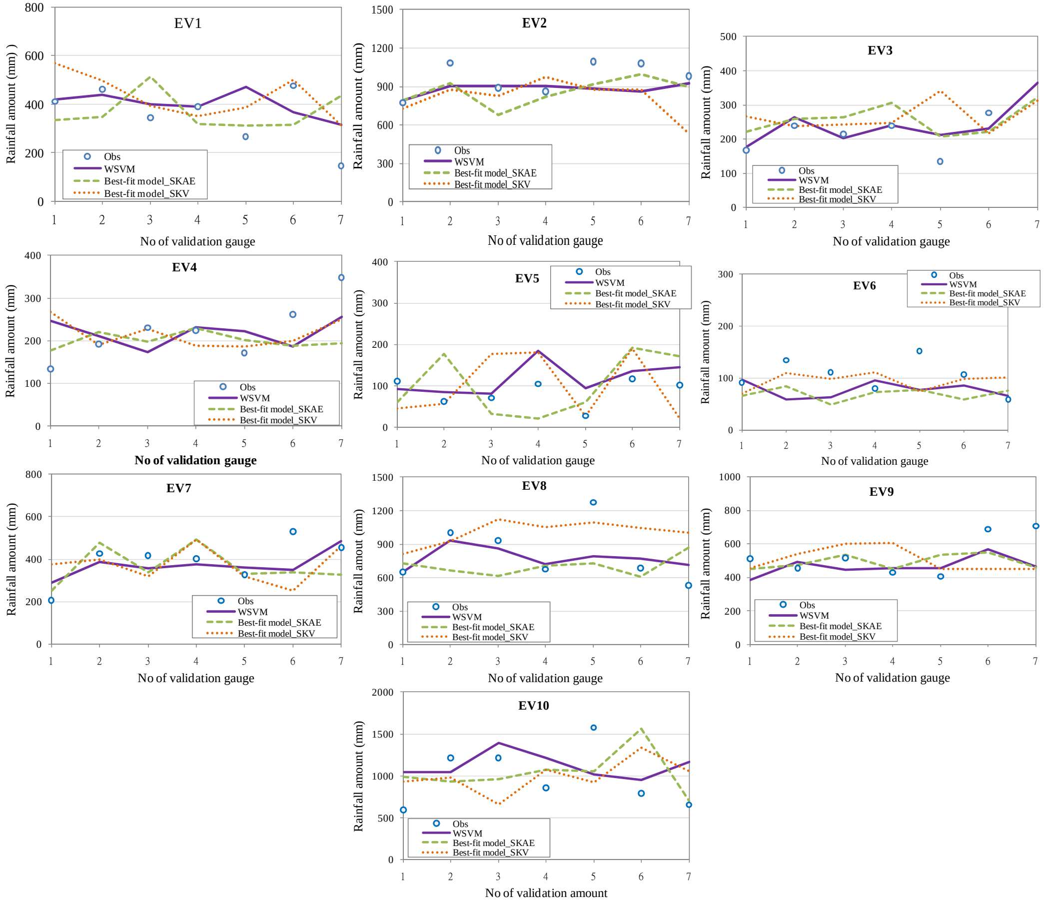

Fig. 3. Comparison of the observed and estimated rainfall amount of different rainstorms by WSVM and best-fit TSVM determined based on SKAE and SKV.

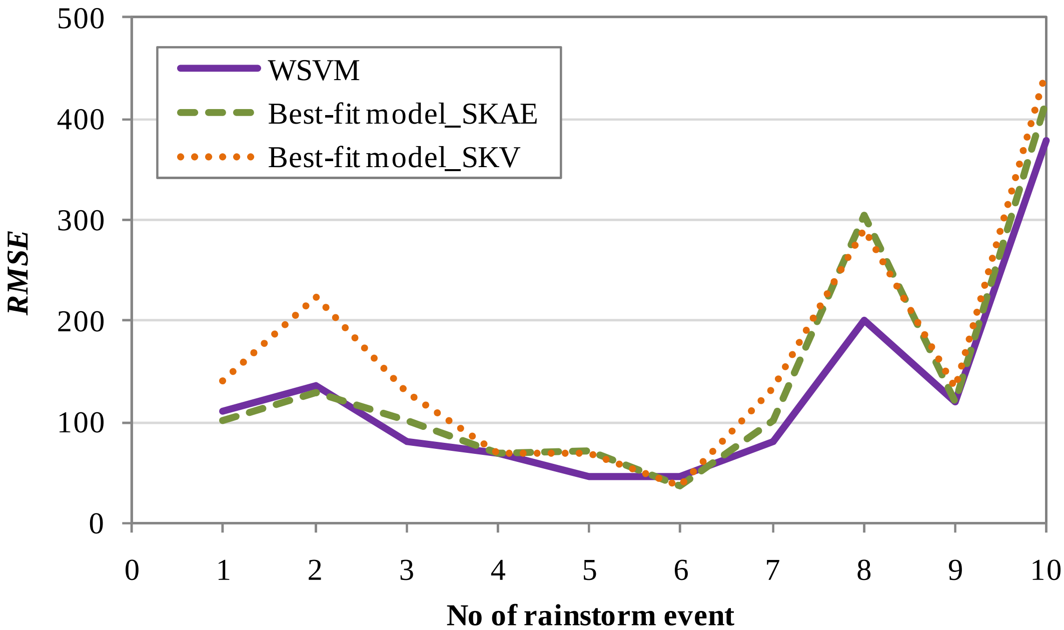

Fig. 4. Root mean square error RMSE of rainfall amount by WSVM and the best-fit TSVM which are identified based on SKAE and SKV.

HESSD
8, 4229–4259, 2011

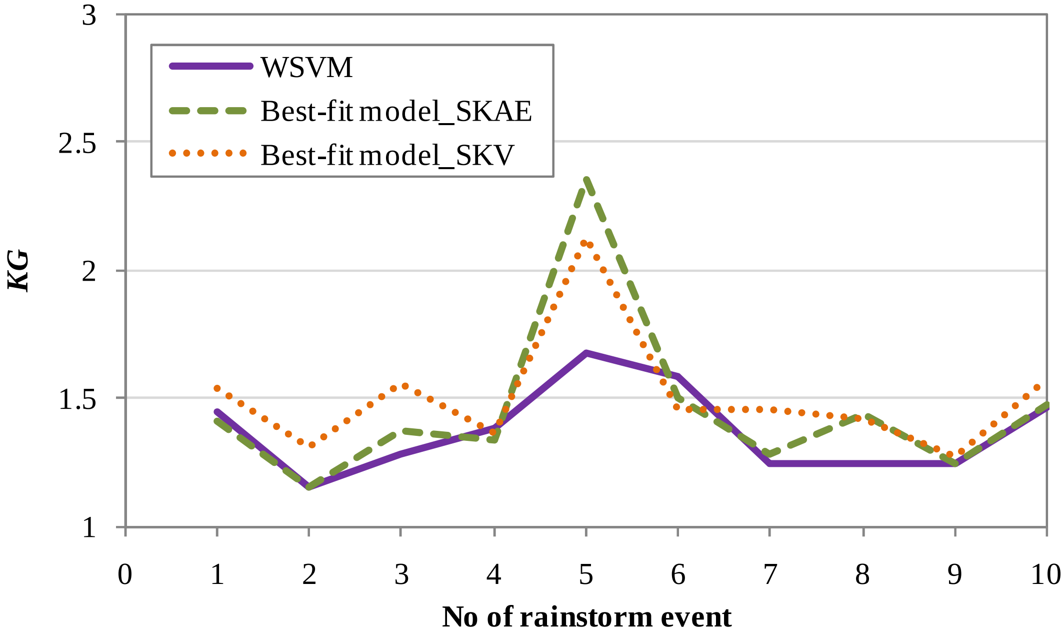

Fig. 5. Model reliability index *KG* of estimated rainfall amount by WSVM and the best-fit TSVM which are identified based on SKAE and SKV.

HESSD
8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.

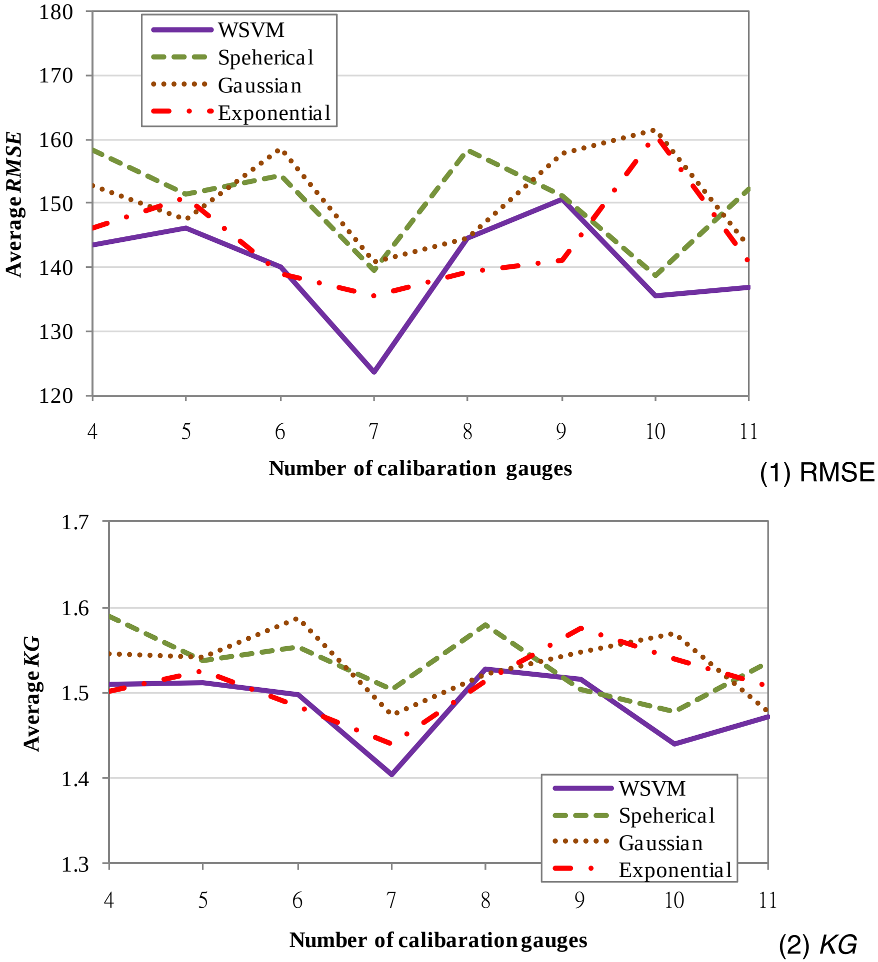

Fig. 6. Comparison of average RMSE and *KG* of estimated amount using WSVM and TSVMs with different number of calibration gauges.

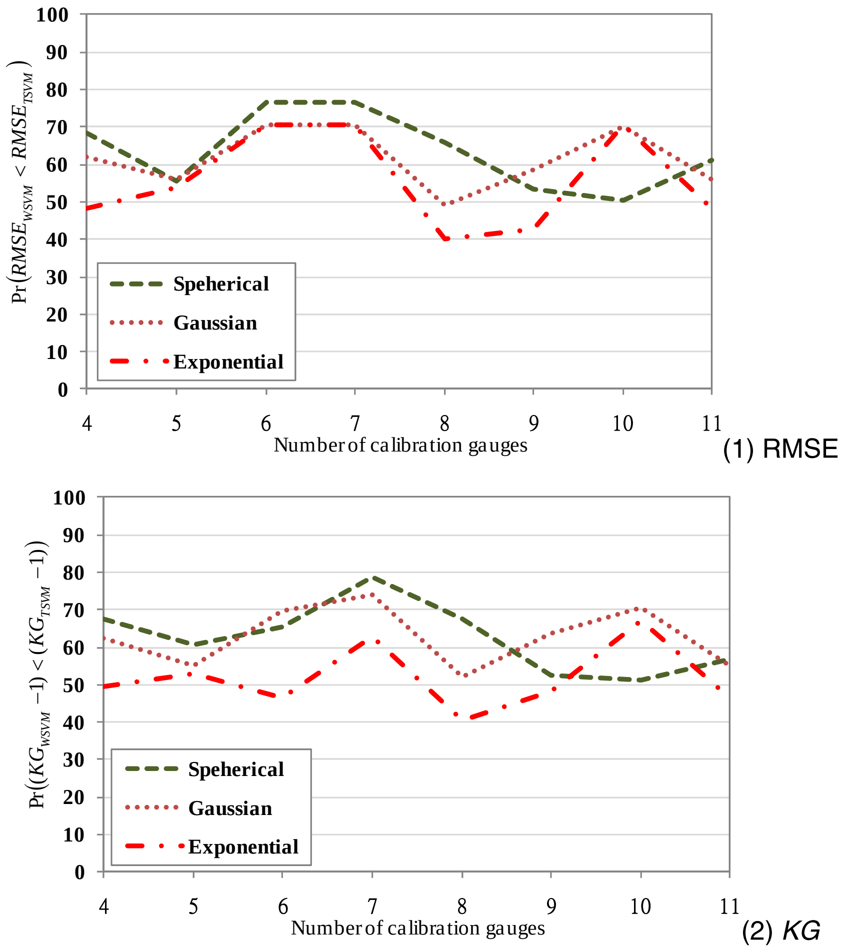

Fig. 7. Probabilities of RMSE and *KG* of estimated rainfall amount using WSVM less than those using TSVM with different number of calibration gauges.

HESSD
8, 4229–4259, 2011

Application of Weighted Semivariogram Model (WSVM)

S.-J. Wu et al.