

Optimal Rain Gauge Network Design and Spatial Precipitation Mapping Based on Geostatistical Analysis from Colocated Elevation and Humidity Data

Aksara Putthividhya and Kenji Tanaka, *Member, IACSIT*

Abstract—The accurate estimation of the spatial rainfall distribution requires a dense network of instruments, which entails large installation and operational costs. It is thus necessary to optimize the number of rainfall stations and estimate point precipitation at unrecorded locations from existing data. This paper serves 2 objectives: i) to establish a spatial representative rainfall stations from the entire existing network in the study area (i.e., rainfall-data optimization); and ii) to use of multivariate geostatistical algorithm for incorporating relatively cheaper hydrological data into the spatial prediction of rainfall. The technique was illustrated using annual and monthly rainfall observations measured at 326 rainfall stations covering Yom river basin and its vicinity in Thailand. Optimal rain gauge network was designed based on the station redundancy and the homogeneity of the rainfall distribution. Digital elevation, humidity, and temperature models were incorporated into the spatial rainfall prediction using multivariate geostatistical algorithms. The results revealed that the multivariate geostatistical algorithm outperform the linear regression, stressing the importance of accounting for spatially dependent rainfall observations in addition to the colocated elevation. The digital elevation data were highly correlated to monthly monsoon-induced precipitation. Humidity and temperature data exhibited a higher degree of correlation to the monthly precipitation data.

Index Terms—Geostatistics, kriging, multivariate, rain gauge network, spatial interpolation.

I. INTRODUCTION

Hydrologic analysis and designs rely significantly on measured rainfall data, including intense storms and flash floods identification based on high resolution estimates of spatial variability in precipitation map. The accurate estimation of the spatial distribution of rainfall usually requires a dense network of instruments, which entails large installation and operational costs. Also, failure of the observer to make necessary visits to the gauge may result in even lower sampling density. It is thus necessary to estimate point rainfall at unrecorded locations from values at surrounding sites.

Manuscript received December 2, 2011; revised March 10, 2012. This work was supported in part by Japan International Cooperation Agency (JICA) under IMPAC-T project. Also, the authors wish to thank “Stimulus Package 2 (SP2) of Ministry of Education under the theme of Green Engineering for Green Society” and “The National Research University Project and the Ratchadaphiseksomphot Endowment Fund (CC513A)” for partial financial support to Aksara Putthividhya.

A. Putthividhya is with the Water Resources Engineering Department, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand (e-mail: dr.aksara.putthividhya@gmail.com).

K. Tanaka is with the Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan (e-mail: tanaka.kenji.6u@kyoto-u.ac.jp).

A number of methods were proposed for the interpolation of rainfall data. The simplest approach, Thiessen polygons, consists of assigning to the unsampled location the record of the closest gauge [1]. Thiessen polygon method was used for estimation of area rainfall [2] as well as applied to the interpolation of point measurements [3]. An inverse square distance method developed by the US National Weather Service in 1972 estimates the unknown rainfall depth as a weighted average of surrounding values (i.e., the weights being reciprocal to the square distances from the unsampled location). It is obvious that both Thiessen polygons and inverse distance methods do not take for other factors such as topography that can affect the rainfall measurement into consideration. The isohyetal method [2] was later developed to use the gauge location and measurement as well as knowledge of the factor affecting these measurements to create isohyets (i.e., lines of equal rainfall). Rainfall at the unsampled location is then estimated by interpolation within the isohyets. This technique also possesses some limitations because an extensive gauge network is usually required to draw isohyets quite accurately.

Geostatistics is based on the theory of regionalized variables [4-6] and allows the spatial correlation between neighboring observations to predict attribute values at unsampled locations. Better geostatistical rainfall estimates (i.e., Kriging) in comparison to the conventional techniques may be achieved [7,8], particularly in sparsely sampled observations. Borga and Vizzaccaro (1997) and Dirks et al. (1998) discovered that for high-resolution networks (e.g., 13 rain gauges over a 35 km² area), the Kriging method does not show significantly greater predictive skill than the inverse square distance method.

Besides providing a measure of prediction error (i.e., Kriging variance), another advantage of Kriging is that the sparsely sampled observations of the primary attribute can be complemented by secondary attributes that are more densely samples [10]. In the case of rainfall, secondary information can be in the form of weather-radar observations [11, 12] and digital elevation [10]. The rationale behind considering digital elevation data, a valuable and cheaper source of secondary information, is that precipitation tends to increase with increasing elevation, mainly because of the orographic effects of mountainous terrain, which causes the air to be lifted vertically, and the condensation occurs due to adiabatic cooling. Multivariate extension of Kriging (i.e., cokriging) as well as Kriging with an external drift [13] were successfully demonstrated to combine both types of information. A more straightforward approach consists of estimating rainfall at a DEM grid cell through a regression of

rainfall versus elevation data [14].

This study aimed to address the issue of incorporating a secondary source of information that tends to be cheaper and more densely sampled into the mapping of precipitation in the Yom river basin area of Thailand. Although estimation for precipitation was demonstrated by accounting for the exhaustive secondary elevation information, the correlation between rainfall and elevation had not yet been applied in Thailand, especially in the study area. Moreover, other secondary information such as humidity and temperature seemed worth accounting into the rainfall mapping since Yom river basin usually experienced both orographic and monsoon rainfall. In this paper, annual and monthly rainfall data from the Yom river basin (Thailand) were interpolated using two types of techniques: (1) univariate methods that use only rainfall data recorded at 326 stations (the inverse square distance and ordinary kriging); and (2) algorithms that combine rainfall data with secondary sources of information including elevation, humidity, and temperature (linear regression and simple kriging with varying local means (SKlm)). Prediction performances of various algorithms were obtained and compared with the traditional inverse distance method, in terms of the pattern of spatial dependence of rainfall.

II. CASE STUDY

A. Study Area

The upper central plain, Thailand, covers about 38,000 km² (180 km × 300 km) of eight provinces with a population of four million people. The main landuse is 63% agricultural, out of which 21% is irrigated, and 24% forest. The basin is surrounded in the East and West by volcanic rocks. The average elevation of the basin is 40-60 m above mean sea level. The basin drains into the lower basin in the South where some free discharge is partially obstructed by crystalline rocks. The 900-1,450 mm annual rainfall within the study region is apportioned to 81% in the wet (Apr-Sept) and 19% in the dry season (Oct-Mar). Fig. 1 shows the location of 326 daily read rain gauge stations (in Yom river basin and its vicinity) employed in this study. The monthly and annual rainfall measurements have been averaged from 1970-2003.

Another source of information is the digital elevation map as shown in Fig. 2. Each grid cell represents 1 km² and its elevation was computed as the average of the elevations at 4 discrete points within the cell. The rationale of incorporating the secondary digital elevation information into the mapping of rainfall in Thailand has been motivated by Goovaerts 2000 [10]. The study showed that the precipitation in Algarve region (Portugal) could be estimated accounting for the exhaustive secondary elevation information. However, the correlation between precipitation and elevation data has not yet been applied in Thailand, especially in the study area. Moreover, other secondary information such as humidity and temperature seem worth accounting into the mapping of monsoon rainfall in Thailand as the strong correlation between precipitation and elevation may be suitable for orographic type of rainfall causing by the vertically lifted air and the condensation from adiabatic cooling [10].

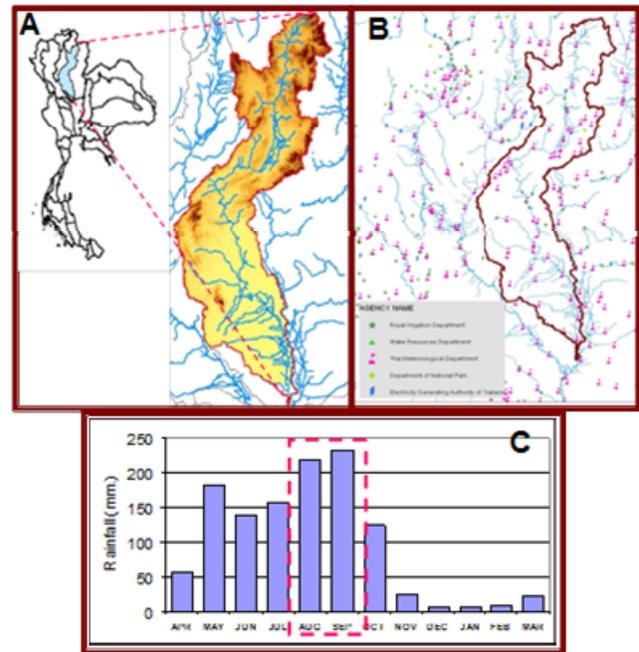


Fig. 1. (a) Study area in Yom river basin; (b) distribution of rain gauges in the study area and vicinity; (c) monthly rainfall at a station in the study area showing the highest rainfall in August and September of Thailand.

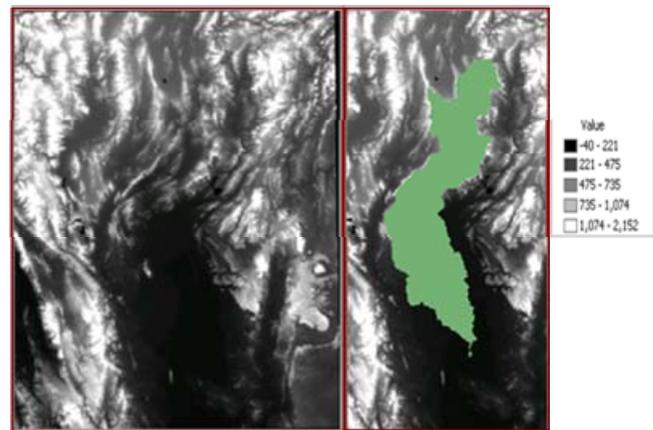


Fig. 2. Digital elevation model of the study area in Yom river basin of Thailand.

III. MAPPING OF PRECIPITATION

The problem is in fact involving the estimation of rainfall depth z at an unsampled location u using rainfall data. Let $\{z(u_\alpha), \alpha = 1, \dots, n\}$ be the set of precipitation values measured at $n = 326$ locations. Elevation is denoted by y and is available at all primary data locations u_α and at all locations u being estimated. Using this information, the precipitation z at any grid node u may be estimated.

A. Inverse Square Distance Method

To avoid unrealistic patchy precipitation maps that could result from Thiessen polygon technique, the rainfall depth z can be estimated as a linear combination of several surrounding observations, with the weights being inversely proportional to the square distance between observations and u . The basic idea behind the weighting scheme is that observations that are close to each other on the ground tend to be more alike than those further apart, hence observations closer to u should receive a larger weight.

$$z_{inv}^*(u) = \frac{1}{\sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)} \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u) z(u_{\alpha}) \quad (1)$$

$$\text{where } \lambda_{\alpha}(u) = \frac{1}{|u - u_{\alpha}|^2}$$

B. Linear Regression

A straightforward approach consists of modeling the relation between elevation and rainfall, e.g., using a linear function of the type:

$$z(u) = f[y(u)] = a_0^* + a_1^* y(u) \quad (2)$$

This relation was then employed to convert each elevation into a rainfall value. A major shortcoming of this linear regression is that the precipitation at a particular grid node u is derived only from the elevation at u regardless of the precipitation value at the surrounding climatic stations u_{α} . Such an approach amounts at assuming that precipitation values are independent from one another, which is rarely the case in practice.

C. Semivariogram

Spatial dependence between observations can be detected using the semivariogram which is a measure of average dissimilarity between observations as a function of the separation vector h . The experimental semivariogram $\hat{\gamma}(h)$ is computed as half the average squared difference between the components of every data pair as follows:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(u) - z(u_{\alpha} + h)]^2 \quad (3)$$

where $N(h)$ is the number of pairs of data locations a vector h apart.

D. Geostatistical Interpolation Approach

Geostatistical interpolation allows us to account for the spatial dependence between observations in the prediction of attribute values. Most of geostatistics is based on the concept of a random function, whereby the set of unknown values is regarded as a set of spatially dependent random variables. Each measurement $z(u_{\omega})$ is thus interpreted as a particular realization of a random variable $Z(u_{\omega})$ [5].

Kriging is a generic name adopted by the geostatisticians for a family of generalized least-squares regression algorithms. The basic idea is to estimate the unknown precipitation value at the unsampled location u as a linear combination of neighboring observations. If only precipitation values are available, the simple Kriging (SK) estimate $z(u)$ is:

$$z_{SK}^*(u) - m = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{SK}(u) [z(u_{\alpha} - m)] \quad (4)$$

where m is the stationary mean of $Z(u)$, and $\lambda_{\alpha}^{SK}(u)$ is the weight assigned to datum $z(u_{\alpha})$ within the search neighborhood $W(u)$. The Kriging weights are determined such as to minimize the estimation variance, while ensuring unbiasedness of the estimator.

E. Simple Kriging with Varying Local Means (SKlm)

The three algorithms introduced herein this work are the variants of simple kriging that allow to incorporate secondary information. Simple kriging with varying local means (SKlm) amounts at replacing the known stationary mean in the SK estimate (4) by known varying means $m_{SK}^*(u)$ derived from the secondary information [5]:

$$z_{SKlm}^*(u) - m_{SK}^*(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{SK}(u) [z(u_{\alpha}) - m_{SK}^*(u_{\alpha})] \quad (5)$$

IV. RESULTS AND DISCUSSIONS

A. Precipitation Map from Univariate Methods

Fig. 3b and Fig. 3c show the map of annual rainfall interpolated at the nodes of $1 \times 1 \text{ km}^2$ grid corresponding to the available resolution of the elevation model using the inverse square distance technique and geostatistical analysis via ordinary kriging differed significantly. Ordinary kriging seemed to provide a smoother precipitation map, while inverse square distance yielded a relatively crude precipitation data. This observation perhaps contributed to nonuniform distribution of the rain gauge network in the study area, stressing the importance of accounting for more densely sampled information.

Another attractive advantage of Geostatistics is that it can easily provide a measure of dissimilarity between observations using the semivariogram. The experimental semivariogram $\hat{\gamma}(h)$, computed from (3), is a function of distance and direction, so it can really account for direction-dependent variability (i.e., anisotropy spatial pattern). The spatial variability was assumed to be identical in all direction, hence only omnidirectional semivariogram was computed in this study. Fig. 4 shows the semivariogram of annual rainfall computed from 326 observations of Fig. 3.

The three models above were fitted using regression and were such that the weighted sum of squares (WSS) of differences between experimental $\hat{\gamma}(h)$ and model $\gamma(h)$ semivariogram values was minimum based on the following equation:

$$WSS = \sum_{\alpha=1}^K \omega(h_{\alpha}) [\hat{\gamma}(h_{\alpha}) - \gamma(h_{\alpha})]^2 \quad (6)$$

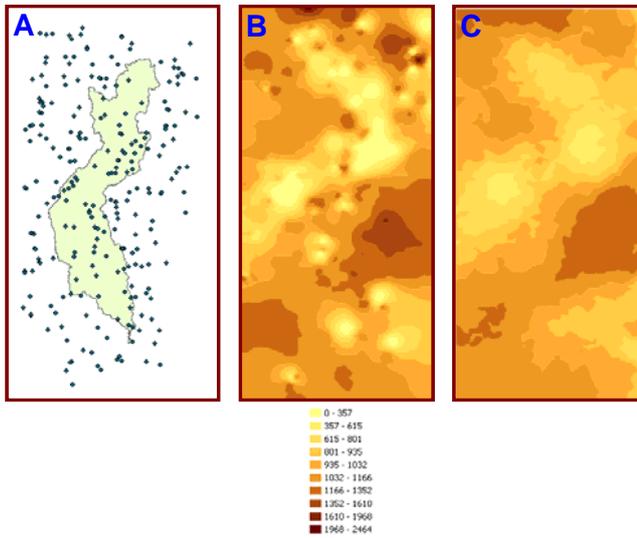


Fig. 3. (a) Distribution of rain gauge in the study area and vicinity; (b) annual rainfall map (mm) obtained by interpolation of 326 observations using inverse square distance method; and (c) annual rainfall map (mm) obtained by interpolation of 326 observations using ordinary Kriging technique.

Calculation of WSS (data not shown) indicated that the spherical model yielded the smallest WSS value. The results from this study agree with others that the spherical model is the most widely used semivariogram model and is characterized by a linear behavior at the origin. Fig. 3C shows the precipitation map generated by ordinary kriging using the spherical semivariogram model.

Semivariogram values increased with the separation distance, reflecting that two rainfall data close to each other are more similar, and thus their squared difference should be even smaller, than those that are further apart. The semivariogram seemed to reach a maximum at 45 km before fluctuating around a still value, supporting the “hole effect” characteristics typically reflected pseudo-periodic or cyclic phenomena [4].

B. Accounting for Elevation

Now, the point of rainfall data ($z(u_\alpha)$) at 326 available stations were supplemented by elevation data ($y(u)$) that are available at all estimation grid nodes. In this particular study, a straightforward approach (i.e. linear regression approach) consists of predicting the precipitation values as a function of co-located elevation according to (2) was employed. Another geostatistical approach, namely simple kriging with varying local means (SKlm) was employed to create representative rainfall map of the Yom river basin in the study area. This replaces the known stationary mean in the simple kriging estimate by known varying means. Since this study consisted of estimating and modeling the semivariogram $\gamma_R(h)$, then SKlm system could be expressed in terms of only covariances as $C_R(0) - \gamma_R(h)$. This was quite convenient as the best semivariogram model was only necessary for precipitation map generation. The precipitation map generated from point estimation of rainfall data at 326 stations supplemented by elevation data using linear regression and SKlm techniques are presenting in Fig. 5.

The results from Fig. 5. revealed that the precipitation map

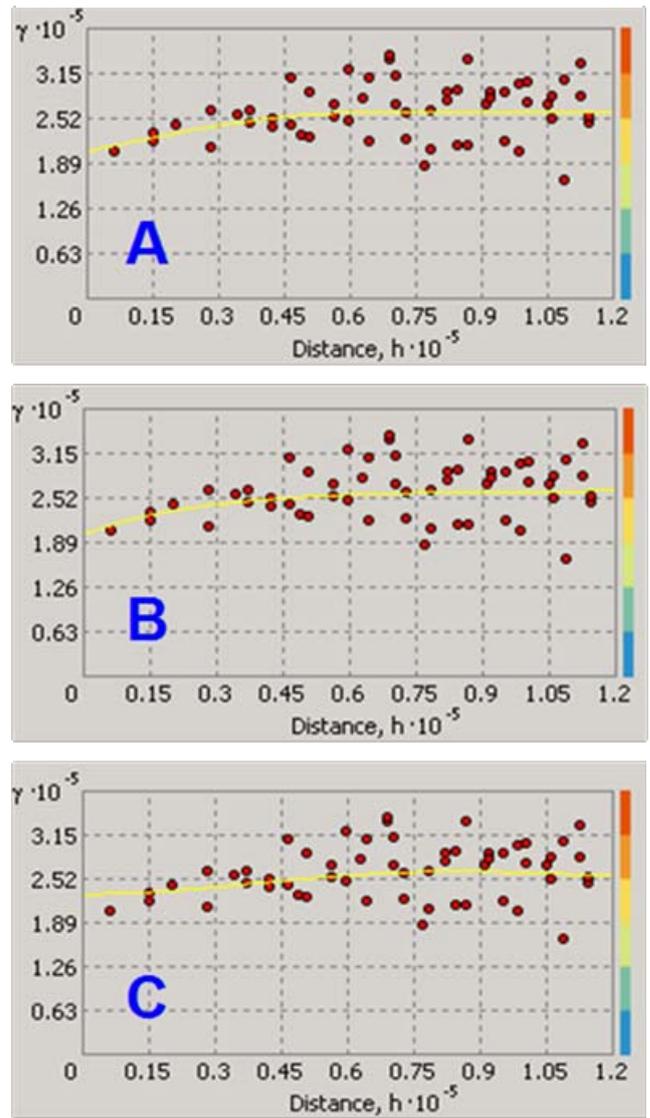


Fig. 4. Experimental semivariogram of the annual rainfall with three different permissible models fitted: (a) spherical model; (b) exponential model; and (c) hole effect model.

generated using SKlm was more similar to the one produced from ordinary kriging approach, indicating that the impact of elevation on rainfall map was less pronounced than for the map generated using linear regression.

The performances of the interpolators described herein this work were assessed and compared using cross validation [15]. Although the various kriging algorithms provided an estimate of the error variance, it could not be retained as a performance criterion because in practice it usually provided little information on the reliability of the kriging estimate [16]. Therefore, the performance of various interpolators were exclusively evaluated. The idea consisted of temporarily removing a set of real rainfall observations from the data set and re-estimate these values from the remaining data using alternate algorithms. The comparison results between the real observations and estimated values could be expressed in terms of mean square error (MSE) of prediction as follows:

$$MSE = \frac{1}{n} \sum_{\alpha=1}^n [z(u_\alpha) - z^*(u_\alpha)]^2 \quad (7)$$

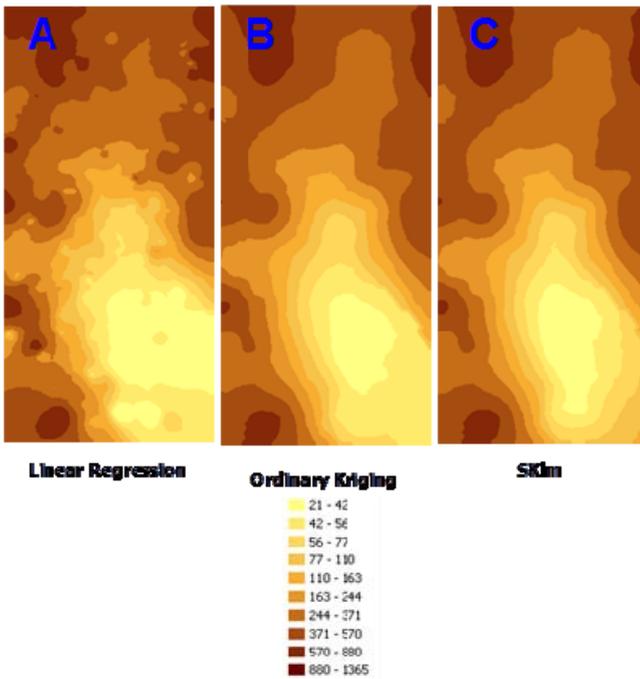


Fig. 5. Annual precipitation maps (mm) obtained by interpolation of 326 observations accounting for the digital elevation model (Fig. 2). Two algorithms were considered: (a) linear regression; (b) simple kriging with varying local means; (c) ordinary kriging.

The algorithms described in the previous sections and illustrated for annual precipitation were also applied to monthly rainfall data. Fig. 6 shows the mean square errors of prediction produced by each geostatistical algorithm employed in this study for the monthly (Jan-Dec) and annual rainfall. Results are expressed as proportions of the prediction error of the linear regression approach.

Relative MSE from SKlm approach shown in Fig. 6 demonstrated the SME values closest to 1.0 in the period of April to September, suggesting that SKlm could perform well for orographic rain in the Yom river basin study area. During the monsoon season (i.e., October-March), the precipitation generated by incorporating elevation data into the precipitation measurements yielded significant errors.

Other secondary information such as humidity seemed worth accounting into the mapping of monsoon rainfall in Thailand as the strong correlation between precipitation and elevation may be suitable for orographic type of rainfall causing by the vertical lifted air and the condensation from adiabatic cooling [10]. Fig. 7 shows the mean square errors if prediction produced by each geostatistical algorithm employed in this study for the monthly (Jan-Dec) and annual rainfall. Results are expressed as proportions of the prediction error of the linear regression approach.

By incorporating humidity data into the mapping of rainfall data in the Yom river basin using SKlm approach, relative monthly SME shown in Fig. 7. demonstrated the SME values closest to 1.0 in the monsoon period (i.e., October-March). The results suggested that the precipitation data generated by incorporating humidity data into the precipitation measurements yielded quite significant improvement in precipitation predictions during the specific period.

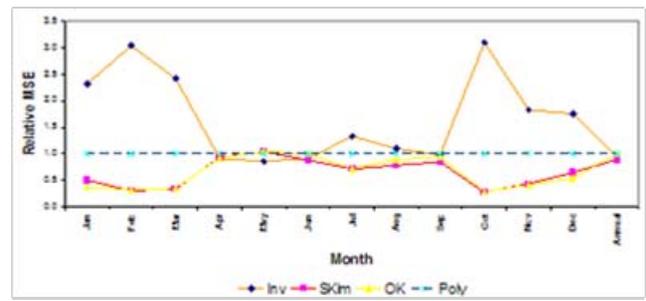


Fig. 6. Relative MSE of prediction produced by each geostatistical algorithm using point precipitation data supplemented by elevation data for the monthly and annual rainfall.

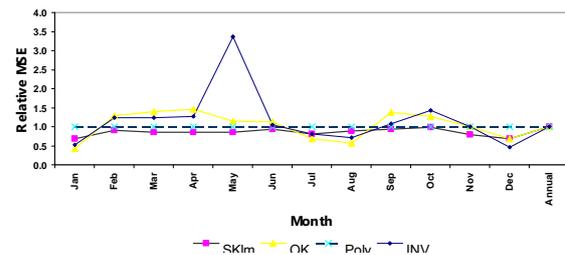


Fig. 7. Relative MSE of prediction produced by each geostatistical algorithm using point precipitation data supplemented by humidity data for the monthly and annual rainfall.

ACKNOWLEDGMENT

The authors thank Mr. Werapol Bejranonda for gathering all necessary information regarding the Yom river basin as he is considered an expert in that specific area. Also, the authors wish to thank Water Resources System Research Unit at Chulalongkorn University (CU_WRSRU) for some hydrologic data used in this research.

REFERENCES

- [1] A. B. Thiessen, "Precipitation averages for large areas," *Monthly Weather Review*, vol. 39, pp. 1082-1084, 1911.
- [2] R. H. McCuen, *Hydrologic Analysis and Design*, Prentice Hall, Englewood Cliffs, N.J., USA.
- [3] K. N. Dirks, J. E. Hay, C. D. Stow, and D. Harris, "High-resolution studies of rainfall on Norfolk island part II: interpolation of rainfall data," *Journal of Hydrology*, vol. 208, pp. 187-193, 1998.
- [4] A. G. Journel, and C. J. Huijbregts, *Mining Geostatistics*, New York, USA: Academic Press, 1978.
- [5] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*, New York, USA: Oxford University Press, 1997.
- [6] P. Goovaerts, "Geostatistics in soil science: state-of-the-art and perspective," *Geoderma*, vol. 89, pp. 1-46, 1999.
- [7] G. Q. Tabios, and J. D. Salas, "A comparative analysis of techniques for spatial interpolation of precipitation," *Water Resources Bulletin*, vol. 21, pp. 365-380, 1985.
- [8] D. L. Phillips, J. Dolph, and D. Marks, "A comparison of geostatistical procedures for spatial analysis of precipitations in mountainous terrain," *Agriculture and Forestry Meteorology*, vol. 58, pp. 119-141, 1992.
- [9] M. Borga, and A. Vizzaccaro, "On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging," *Journal of Hydrology*, vol. 195, pp. 160-171, 1997.
- [10] P. Goovaerts, "Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall," *Journal of Hydrology*, vol. 228, pp. 113-129, 2000.
- [11] J. D. Creutin, G. Delrieu, and T. Lebel, "Rain measurement by raingage-radar combination: a geostatistical approach," *Journal of Atmospheric and oceanic technology*, vol. 5, pp. 102-115, 1988.

- [12] A. Azimi-Zonooz, W. F. Krajewski, D. S. Bowles, and D. J. Seo, "Spatial rainfall estimation by linear and non-linear cokriging of rainfall and raingage data," *Stochastic Hydrology and Hydraulics*, vol. 3, pp. 51-67, 1989.
- [13] G. Raspa, M. Tucci, and R. Bruno, "Reconstruction of Rainfall Fields by Combining Ground Raingauges Data with Radar Maps using External Drift Method," in *Geostatistics Wollongong '96*, E. Y. Baafi, and N.A. Schofield, Eds. Dordrecht, Kluwer Academic, 1997, pp. 941-950.
- [14] C. Daly, R. P. Neilson, and D. L. Phillips, "A statistical topographic model for mapping climatological precipitation over mountainous terrain," *Journal of applied meteorology*, vol. 33, pp. 140-158, 1994.
- [15] E. H. Isaaks, and R. M. Srivastava, *An Introduction to Applied Geostatistics*, New York, USA, Oxford University Press, 1989.
- [16] M. Armstrong, "Is Research in Mining Geostatistics as Dead as a Dodo?" in *Geostatistics for the Next Century*, R. Dimitrakopoulos, Ed. Dordrecht, Kluwer Academic, 1997, pp. 303-312.

Aksara Putthividhya was born in Bangkok, Thailand, in 1977. She got a M.S. and Ph.D. in Environmental and Water Resources Engineering from The University of Michigan, Ann Arbor, USA in 2000 and 2004, respectively. Now, s Next, the author's educational background is

She is now working as an Assistant Professor at the Department of Water Resources Engineering, Chulalongkorn University. Her current research interests are related to groundwater contamination, fate and transport of organic pollutants and microbes in subsurface formation, flood modeling, probabilistic flood discharge estimation, integrated water resources management modeling, and stable isotope techniques in investigating surface water-groundwater interactions. She has received several research fellowships from Thai governments and recently awarded a research funding related to flood from JICA.

Dr. Putthividhya is a member of AGU, JPGU, AOGS, IAH, and HAT (IAH Thailand).