JOURNAL OF GEOGRAPHIC INFORMATION AND DECISION ANALYSIS

Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 243 - 251, 1998

A Methodological Study of the Application of the Maximum Entropy Estimator to Spatial Interpolation

Yuh-Ming Lee

Institute of Natural Resource Management, National Chung Hsing University 69, Chienkwo N. Rd., Sec. 2, Taipei 104 Taiwan R.O.C.

yml@cc.nchulc.edu.tw

ABSTRACT **This study applies the maximum entropy estimator to the interpolation of daily rainfall measurements in Switzerland taken on April 26, 1986. The major purpose of the paper is to introduce the methodology of maximum-entropy spatial interpolation. The accuracy of the estimation measured by the coefficient of determination is therefore only at the value of 0.0114. Moreover, the estimation errors are, to some extent, correlated to the real measurements. The larger the measurement is, the more severe the error becomes. The structure analysis on the 100 measurements used for estimation depicts the properties of anisotropy and non-stationarity in spite of the assumptions of second-order stationarity and isotropic correlation. The estimation errors, however, seem to be spatially independent. The spatial correlation coefficients of the errors (calculated based on the construction of the correlogram of the estimation errors) are only at the order of 0.01.**

KEYWORDS: **Shannon's entropy, maximum entropy principle, lognormal random field, simple kriging, log-kriging.**

Contents

1. Introduction

This paper is principally the result report of the participation of SIC '97. The focus and purpose of the study are a methodological introduction of the maximum entropy estimator (M.E.E. hereafter) rather than the development of an effective interpolator. The estimation results, thus, may pose some errors or biases. The paper consists of, in addition to the brief introduction, four parts: the description of the M.E.E., implementation and assumptions, results and discussion, and the conclusion. The data used are the rainfall measurements in Switzerland distributed by the editorial board. Some descriptive statistics and histograms are shown as Figure 1 and Table 1 (to be described in the section of interpolation results and discussion). The histograms show that the rainfall data seem to be log normally distributed rather than the commonly assumed normal distribution.

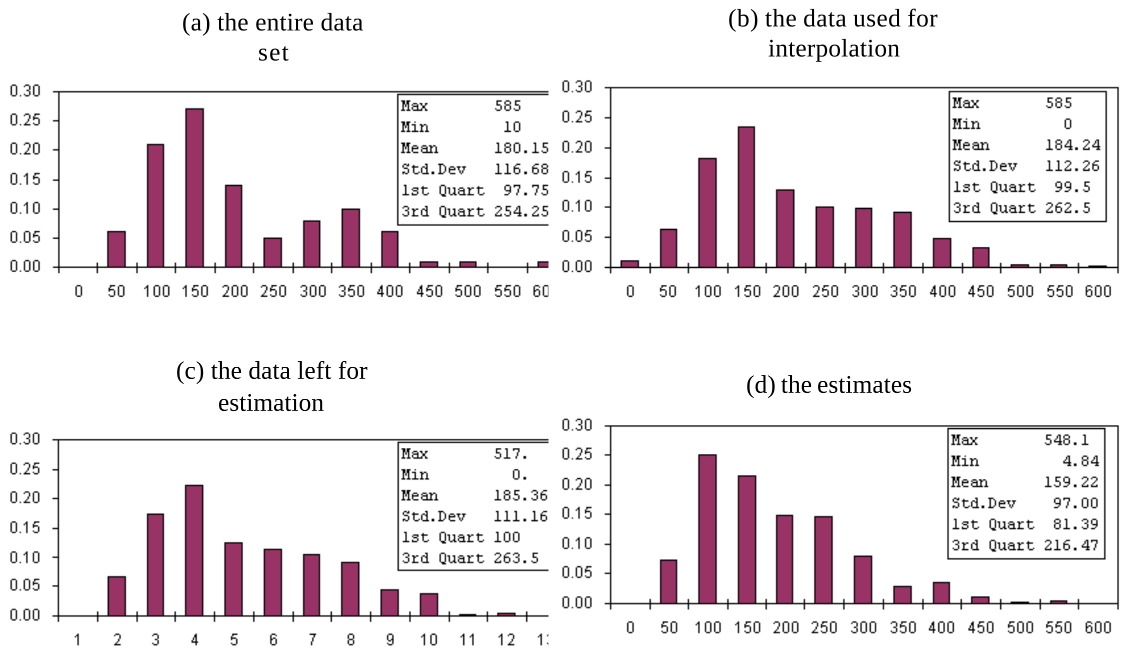

Figure1 Histograms and descriptive statistics of the data set, data used for interpolations, data left for estimation, and the estimates

2. The Maximum Entropy Estimator

The maximum entropy estimator is based on Shannon's (information) entropy concept (Shannon, 1948) and the maximum entropy principle (Jaynes, 1957). Given expectations (*i.e.*, prior information) of a univariate random variable or a multivariate random function, the maximum entropy principle can be used to derive a minimally prejudiced (*i.e.*, least presumptive) probability distribution (Tribus, 1978; Jaynes, 1982; Theil and Fiebig, 1984; Kapur, 1989). For a multivariate continuous random function, this principle can be described by an optimization problem, Maximize:

$$h(Z) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(z_1, z_2, \cdots, z_m)\cdot \log f(z_1, z_2, \cdots, z_m)\, dz_1 \cdots dz_m \tag{1}$$

Subject to:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(z_1, z_2, \cdots, z_m)\cdot dz_1 \cdots dz_m = 1 \tag{2}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(z_1, z_2, \cdots, z_m)\cdot g_r(z_1, z_2, \cdots z_m)\, dz_1 \cdots dz_m = \hat{g}_r; \quad r = 1, 2, \cdots, q \tag{3}$$

where $f(z_1, z_2, \cdots, z_m) = f(\mathbf{z})$ is the probability density function (pdf) of $\mathbf{Z} = (Z_1, Z_2, \cdots Z_m)$; $\varepsilon(\mathbf{Z}) = \log f(\mathbf{z})$ denotes Shannon information; $\hat{g}_r$ denotes expectations of the functions $g_r(\mathbf{z})$ with respect to the random vector $\mathbf{Z}$.

The maximum entropy estimator of $Z_0 = Z(\mathbf{x}_0)$ minimizes the posterior uncertainty (conditional information) of Z_0 given $\mathbf{z}_g = (z_1, z_2, \cdots z_n)$. Given a set of data locations $(\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_n)$ and the measurements $\mathbf{z}_g = (z_1, z_2, \cdots z_n)$, the conditional information, denoted $\varepsilon(Z_0 \mid \mathbf{z}_g)$, is minimal with respect to the optimal (maximum entropy) estimate of $\hat{Z}(\mathbf{x}_0) = \hat{z}_0^*$. The solution to the following formulation (necessary condition for a minimum) determines the maximum entropy estimator:

$$\left.\frac{d\varepsilon(Z_0 \mid \mathbf{z}_g)}{dz_0}\right|_{z_0=\hat{z}_0^*} = \left.\frac{d}{dz_0}\log f(z_0 \mid \mathbf{z}_g)\right|_{z_0=\hat{z}_0^*} = \left.\frac{d}{dz_0}\log\frac{f(z_0, \mathbf{z}_g)}{f(\mathbf{z}_g)}\right|_{z_0=\hat{z}_0^*} = 0 \tag{4}$$

where $f(z_0 \mid \mathbf{z}_g)$ is pdf of Z_0 given $\mathbf{z}_g$; and, $f(z_0, \mathbf{z}_g)$ is the joint pdf. The derivative of $\log f(\mathbf{z}_g)$ with respect to $z_0 = \hat{z}_0^*$ is zero because the prior pdf $f(\mathbf{z}_g)$ is independent of $\hat{z}_0^*$. Equation (4) then becomes:

$$\left.\frac{d\log[f(z_0, z_1, \cdots z_n)]}{dz}\right|_{z_0=\hat{z}_0^*} = 0 \tag{5}$$

The above equation is the necessary condition for a minimum. Given concavity of the logarithmic function (or convexity of the information function) the condition is also sufficient.

For the prior information with the first moments given (*i.e.*, a Gaussian random field), the maximum entropy estimate of Z_0 given $\mathbf{z}_g = (z_1, z_2, \cdots z_n)$ is (Lee and Ellis, 1997):

$$\hat{Z}_0^* = m_0 + (\Sigma_{gu}^0)^t \Sigma_{gg}^{-1} (\mathbf{z}_g - \mathbf{m}_g) \tag{6}$$

where $\mathbf{m}_g, \Sigma_{gu}^0, \Sigma_{gg}$ are the given first two moments. The estimation error associated with the estimate is:

$$err(\hat{Z}_0^*) = Var(\hat{Z}_0^* - Z_0) = \sigma_0^2 - (\Sigma_{gu}^0)^t \Sigma_{gg}^{-1} (\Sigma_{gu}^0) \tag{7}$$

where σ_0^2 is the variance of Z_0 (log-transformed variates). The maximum entropy estimator and the associate estimation error of a lognormal field (with the means and variances of the log-transformed variates given) are (Lee and Ellis, 1997):

$$\hat{Y}_0^* = \exp\left(\hat{Z}_0^* \quad err(\hat{Z}_0^*)\right) \tag{8}$$

$$err(\hat{Y}_0^*) = e^{2\mu} \cdot err(\hat{Z}_0^*) \tag{9}$$

where $Z_i = \log Y_i$ is the log-transformed variables; e^{μ} is the geometric mean; $\hat{Z}_0^*$ and $err(\hat{Z}_0^*)$ are the maximum-entropy estimates and the error variances of the log-transformed variates which are normally distributed. The lognormal M.E.E. is not an exact estimator. The biases of the estimates at sampled locations are: $\{\exp(-\frac{3}{2}\sigma_i^2); \forall i\}$.

3. Assumptions and Implementation

The focus of the study is to introduce methodologically the maximum entropy estimator. Thus, we simplify the procedure of data analysis. For both the presumed Gaussian and lognormal fields, we assume constant means and isotropic covariance structures despite they seem to be anisotropic. The well-known second-order stationarity and isotropic correlation are assumed first to apply the maximum entropy estimator. These yield a Gaussian random field when the principle of maximum entropy is applied. Unfortunately, some of the estimates yield negative values that depart from the physical meanings. The lognormal maximum entropy estimator is therefore applied. The underlying assumptions of the lognormal maximum entropy estimator are a second-order stationary and an isotropic correlation structure on the log-transformed variates. Based on structure analysis of the precipitation data, the correlation structure can be described fairly by a Gaussian model of:

$$\mathrm{Cov}\left(Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{h})\right) = C(|\boldsymbol{h}|) = 0.5724 \cdot \exp\left(-\frac{3 \cdot |\boldsymbol{h}|^2}{35000^2}\right) \tag{10}$$

where 0.5724 is the sample variance calculated from the log-transform values of the 100 observations; 35000 (m) is the so-called "range". The experimental covariance structure and the model are illustrated as Figure 2.

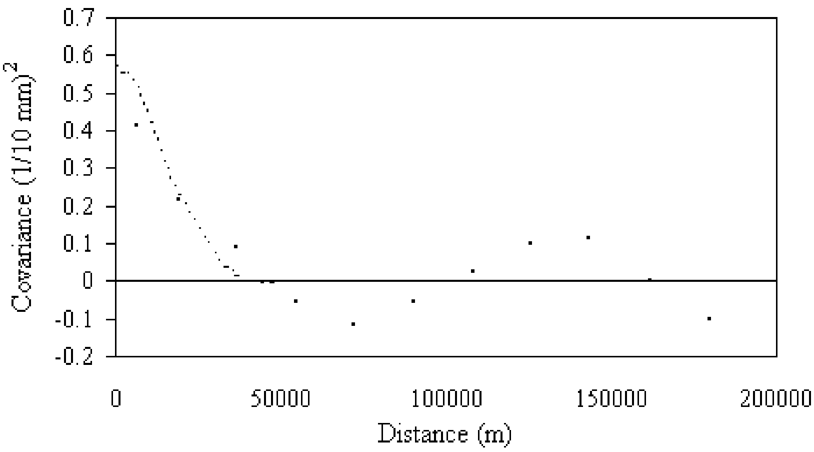

Figure 2 Correlation structure under assumption of isotropy and second-order stationarity on the log-transformed variates

The maximum entropy estimator is still at the development stage. There seem to be no commercial programs or codes available. A Fortran 77 program is therefore coded for use in the study. The program takes the covariance structure as a user-defined function. Measurements and locations to be estimated are read in with conventional program input/output procedure. Figures and maps are constructed by using the Microsoft Excel and Surfer (Golden Software).

4. Interpolation Results and Discussions

The estimator formulated as Equation (8) is implemented as a Fortran 77 program developed by the author. The estimation results have been submitted to the editorial board. Some descriptive statistics of the estimates associated with those of the entire data set and the data used for estimation are listed in Table 1. The underlying assumption of the M.E.E. applied is a lognormal random field. The mean and variance of the data and estimates are also calculated based on the assumption of lognormal distribution (denoted as LN-Mean and LN-Variance). The estimation errors with proportional symbols denote the differences between estimates and measurements and are plotted as Figure 3. We define here the estimation error, rather than the error variances described previously, as the difference between the estimates and the measurements. The isoline maps of estimates associated with the lowest / highest 10 measurement locations are plotted as Figure 4. As shown in Figure 4, the M.E.E. identifies 4 of the 10 lowest / highest measurements respectively. The characteristics of estimation error, such as descriptive statistics and correlation, are listed as Table 2 The bias of errors (*i.e.*, the mean of absolute errors) is 70.62. Moreover, the errors seem strongly correlated to the real measurements. The correlation coefficient between errors and measurement is -0.5604 for the errors and 0.4975 for the absolute errors.

Table 1 Statistical characteristics of the data set, the data used, and the estimates			
Statistical Properties	All Data	Data Used	Estimates by M.E.E.
Minimum	0	10	4.84
Maximum	585	585	548.10
Mean	184.24	180.15	159.22
Median	152	141	138.84
Standard Deviation	112.26	116.68	97.00
Geometric Mean*	148.44	142.02	129.58
1st Quartile	99.50	97.75	81.39
3rd Quartile	262.50	254.25	216.47
Log-Mean**	5.00	4.96	4.86
Log-Variance**	0.85	0.58	0.48
LN-Mean***	227.16	189.63	164.96
LN-Variance***	120853.31	64112.41	44103.68

* Measurements with values of zero are omitted when calculating geometric means.

** Log-Mean and Log-Variance denote mean and variance of log-transformed values of measurement respectively.

*** LN-Mean and LN -Variance are the mean and variance of the data under the assumption of the lognormal distribution.

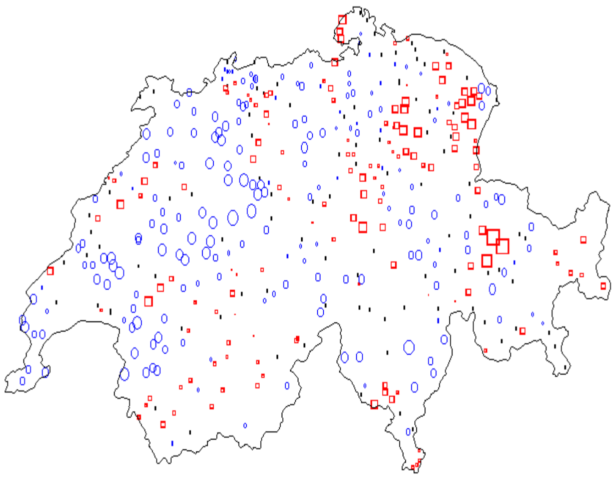

Figure 3 Estimation errors with proportional symbols: □ -- overestimated; ○ -- underestimated; ● -- data used for estimation

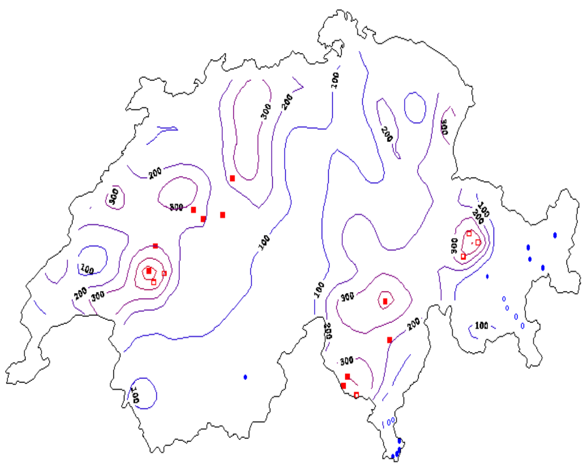

Figure 4 Isoline maps of the estimates in which the method of triangulation with linear interpolation is used for drawing the isolines. The locations associated with the minimal and maximal ten measurements / estimates are posted as circles and squaresrespectively. The solid symbols denote those of the measurements.

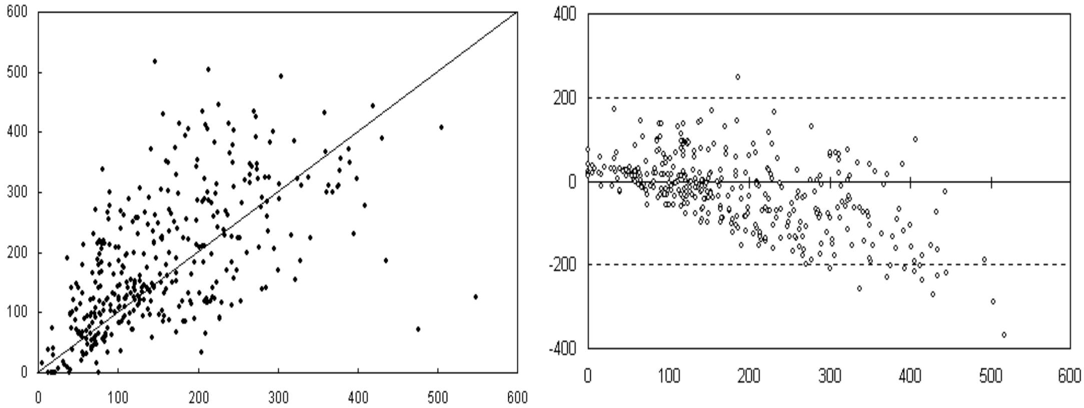

Figure 5 Scatter plots of (a) estimates vs. measurements and (b) error vs. measurements where the measurements are plotted as the horizontal axis.

Table 2 Statistical description of the estimation errors

* Errors are defined as the difference between estimated values and measurements; ** Measurements with the value of zero are omitted.

	Errors*	Absolute Errors
Mean	-26.14	70.62
Standard Deviation	92.83	65.59
Minimum	-370.55	0.03
Maximum	422.10	422.10
Correlation with Data	-0.5604	0.4975
Correlation with MEE	0.3147	0.2668
Mean Squared Errors (MSE)		9327.50
Root Mean Squared Errors (RMSE)		96.58
Error Sum of Squares (ESS)		3404536.74
Total Sum of Squares (TSS)		3443709.41
Mean Absolute Errors		70.62
Mean Relative Errors**		0.05132
Coefficient of Determination R^2		0.01138

The descriptions of estimation accuracy here are most likely similar to those of a regression problem in which the slope is expected to be one and the intercept should be zero. Some measures of goodness-of-fit used in econometrics are therefore evaluated and listed in Table 2. The coefficient of determination R^2 is selected as the measure of the accuracy of the estimation. The coefficient of R^2 is only 0.01138 which is far from one (perfect fitt). For further comparison, two scatter plots showing the correlation between estimates / errors and measurements are presented as Figure 5. The estimates seem to be mostly underestimated and are to some extent, correlated to the measurements. The larger the measurements are, the more severe the estimation errors become. An extended structure analysis on the correlogram of the estimation errors is carried out. The sample correlation coefficients are only at the orders of 0.001 to 0.01, which are far from unity and no decay-type structure can be observed. The errors, therefore, appear to be relatively least spatially correlated. Moreover, in the case of an emergency such as radioactivity monitoring in a nuclear accident, the M.E.E seems more applicable in an automated system for quick interpolation of spatial measures, mainly due to its similarity to simple log-Kriging and fast computing. The M.E.E. does not need intensive pre-modeling nor a great amount of data analysis.

5. Summary and Conclusion

This study applies the maximum entropy principle to spatial interpolation. The method is called the maximum entropy estimator. The major purpose of the paper is to introduce the still-developing methodology. The estimates are therefore not highly accurate. Nevertheless, the study investigates more closely the errors produced by the maximum entropy estimator. The estimation errors are, to some extent, correlated to the real measurements. The larger the measurement is the more severe the error becomes. Moreover, the errors seem to be spatially independent.

The maximum entropy estimator introduced in this study applies the definition of Shannon's entropy of continuous random variables. The formalism of maximum entropy is, therefore, in the form of integral. With a large number of measurements, the discrete form of Shannon's entropy evaluations (*i.e.*, summation instead of integral) is more applicable. The recently developed definition of spatial entropy or spatial disorder by Journel and Deutsch (1993) seems to pose great potential for further development of the discrete or block forms of the maximum entropy estimator.

References

Christakos, G. (1990) A Bayesian/Maximum-Entropy View to the Spatial Estimation Problem, *Mathematical. Geology*, 22(7), 763-777.

Christakos, G. (1991) Some Applications of the Bayesian/Maximum-Entropy Concepts in Geostatistics, in *Maximum Entropy and Bayesian Methods*, 215-229, Grandy, W. T. and Schick, L. H., eds., Kluwer Academic Publishers.

Christakos, G. (1992) *Random Field Models in Earth Sciences*, San Diego, CA: Academic Press.

Jaynes, E. T. (1957) Information Theory and Statistical Mechanics, *Physical Review*, 106(4), 620-630.

Jaynes, E. T. (1982) On the Rationale of Maximum-Entropy Methods, *Proceedings of the IEEE*, 70(9), 939-952.

Journel, A. G. and C. V. Deutsch (1993) Entropy and Spatial Disorder, *Mathematical Geology*, 25(3), 329-355.

Kapur, J. N. (1989) *Maximum-Entropy Models in Science and Engineering*, New York: John Wiley & Sons.

Lee, Y.-M. and J. H. Ellis (1997) Estimation and Simulation of Lognormal Random Fields, *Computers and Geosciences*, 23(1), 19-31.

Lee, Y.-M. and J. H. Ellis (1997) On the Equivalence of Kriging and Maximum Entropy Estimators, *Journal of Mathematical Geology*, 29(1), 131-152.

Shannon, C. E. (1948) A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27, 379-423, 623-656. (Reprinted in *Claude Elwood Shannon Collected Papers*, Sloane, N. J. and Wyner, A. D., Eds., IEEE Press, pp. 5-83.1993).

Theil, H. and Fiebig, D. G. (1984) *Exploiting Continuity, Maximum Entropy Estimation of Continuous Distributions*, Cambridge, MA: Ballinger Publisher.

Tribus, M. (1978) Thirty Years of Information Theory, in *The Maximum Entropy Formalism* (Levine, R. D. and Tribus, M., eds.), Cambridge, MA: The MIT Press.

JGIDA vol. 2, no. 2
JGIDA Home