韓國水資源學會論文集
第32卷 第1號 · 1999年 2月
pp. 3~13

신경망 기법을 이용한 연평균 강우량의 공간 해석

Spatial Analysis for Mean Annual Precipitation Based On Neural Networks

신 현 석* / 박 무 종**
Shin, Hyun Suk / Park, Moo Jong

Abstract

In this study, an alternative spatial analysis method against conventional methods such as Thiessen method, Inverse Distance method, and Kriging method, named Spatial-Analysis Neural-Network (SANN) is presented. It is based on neural network modeling and provides a nonparametric mean estimator and also estimators of high order statistics such as standard deviation and skewness. In addition, it provides a decision-making tool including an estimator of posterior probability that a spatial variable at a given point will belong to various classes representing the severity of the problem of interest and a Bayesian classifier to define the boundaries of subregions belonging to the classes. In this paper, the SANN is implemented to be used for analyzing a mean annual precipitation filed and classifying the field into dry, normal, and wet subregions. For an example, the whole area of South Korea with 39 precipitation sites is applied. Then, several useful results related with the spatial variability of mean annual precipitation on South Korea were obtained such as interpolated field, standard deviation field, and probability maps. In addition, the whole South Korea was classified with dry, normal, and wet regions.
Keywords: Spatial Analysis, Mean Annual Precipitation, Neural Networks

요 지

본 연구에서는 공간 분포의 해석을 위하여 일반적으로 사용되어 오던 Thiessen 또는 Kriging 법들을 대체할 수 있는 새로운 모형인 SANN(Spatial-Analysis Neural-Network)을 소개한다. 이 모델은 신경망 기법을 이용한 비매개변수법의 일종으로 미측정 지점의 평균값 뿐만 아니라 분산, 왜도 등의 고차 통계치를 제공하여 준다. 또한 어떤 지점에서의 공간변수의 값이 그 심각도에 따른 미리 지정된 여러 분류들 중 각각의 분류에 속할 확률값과 전체 공간을 각 분류에 따라 가장 최적하게 분류경계(class boundary)를 선정하여 줄 수 있는 Bayesian 계급분류기(Classifier)를 제공하는 의사결정(decision-making) 역할도 수행할 수 있다. 본 연구에서는 제안된 SANN모형의 외삽기(interpolator)를 사용하여 관측 지점의 연평균 강우량을 대상 유역 전체에 공간적으로 분포시키고 또한 각 지점의 예측 오류를 산정하며, Bayesian 분류기를 사용하여 대상유역을 가장 적절하게 건조, 보통, 습윤 지역으로 분류하는 방법을 제시하여 본다. 본 연구에서는 39개 강우 계측 지점을 이용하여 우리나라의 연평균 강우의 공간 해석에 응용하여 본다. 결과적으로 연평균 강우량의 공간 분포, 표준편차, 그리고 확률도를 얻었다. 더불어 우리나라 전역을 건조, 보통, 습윤 지역으로 분류하여 보았다.
핵심용어 : 공간분석, 연평균 강수량, 신경망

* 부산대학교 토목공학과 조교수
Assistant Prof., Dep. of Civil Engrg., Pusan National Univ., Pusan 609 735, Korea
** 한서대학교 토목공학과 조교수
Assistant Prof., Dep. of Civil Engrg., Hanseo Univ., Chungnam 356 820 Korea

1. Introduction

The uncertainty of hydrological and environmental data such as precipitation, soil properties, and groundwater contaminant concentration is of a great necessity and importance for solving various problems related to water resources planning and management, groundwater contamination, and water quality control. A number of methods such as Kriging have been suggested in literature for hydrological and environmental data (Bras and Rodrigues-Iturbe, 1985). However, these methods have been limited for analyzing the complex natural phenomena, because of the assumptions of stationarity and normality of the underlying variables and the drawbacks in structural analysis such as shadow effect, anisotrophic data, nested structure, and hole effect (ASCE, 1990).

Recently, Neural Networks have been successfully used to solve some complex hydrological and environmental problems such as river flow prediction (Markus et al., 1995), activated sludge prediction (Novotny et al., 1991), determination of aquifer parameters (Rashid et al., 1992), and groundwater reclamation (Rogers and Dowla, 1994). Based on various advantages of the use of neural networks (Haykin, 1994), Shin and Salas (1997) introduced a alternative method, called Spatial Analysis Neural Network (SANN) which pertains the following characteristics: (1) nonparametric estimators of the conditional mean and higher order moments such as standard deviation and skewness coefficient; (2) the estimator of the point posterior probability estimator for some classes predefined and the Bayesian classifier to assign a class to an arbitrary spatial point. The proposed estimators are implemented into a specified multi layer feed-forward neural network structure to achieve computational efficiency based on parallel system modeling, and its structure and operation scheme is summarized briefly in this paper.

Mean Annual Precipitation (MAP) for a given region is commonly required for several hydrological studies such as water-balance calculations, groundwater flow modeling, and drought area investigation. For these studies, it is necessary to estimate MAP at arbitrary stations using the observed precipitation data at the gauged stations. Tabios and Salas (1985) compared several methods for estimating MAP and concluded that Ordinary Kriging (OK) and Universal Kriging (UK) were superior to Thiessen polygon, ID, and PN methods. Likewise, Kassim and Kottegoda (1991) compared OK and Disjunctive Kriging (DK) for designing the MAP network.

In this paper, there are two purposes. The first is to introduce the Spatial Analysis Neural Network (SANN) for analyzing the spatial variability of the MAP of the whole area of South Korea. Based on this application, the applicability of SANN may be verified for several other applications associated with the analysis of spatial variability. The other is for the region to be classified into dry, normal, and wet precipitation areas, which can not be accomplished by the traditional methods such as Kriging. For doing this, the 39 precipitation sites where the observation periods are long comparably were used as the observation precipitation sites.

2. Description of Spatial Analysis Neural Network (SANN) Model

In this section, we describe the structure and operation of SANN, which has developed for analyzing any type of spatial variables, based on a multi-layer feed-forward neural network form. The more theoretical derivations and illustrations related to various estimators can be found in Shin and Salas (1997).

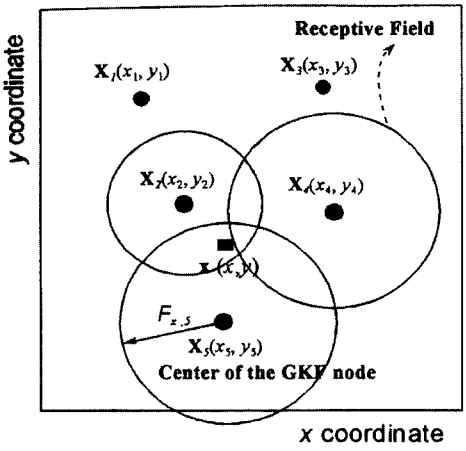

Fig. 1. Arbitrary spatial region in a two-dimensional domain. The circular boundary represents the receptive field of the Gaussian Kernel Function (GKF) node, $\sigma_{x,n}$ represents the width of the Gaussian kernel at point n, and $X_n(x_n,y_n)$ is the center of the n-th GKF node. x(x,y) is any interpolation point.

Suppose that we configure a spatial region $\mathcal{R}$ as shown in Fig. 1. Assume that measurements of the spatial variable, z, are available in a two-dimensional domain, i.e. $\mathbf{x} = [x, y]$. We have N sample observations in the region which are denoted by the observation set $\{ \mathbf{X}_n, Z_n \mid n = 1, ..., N\}$. We want to determine the values of z at any point $\mathbf{x}$, *i.e.* $\hat{z}(\mathbf{x})$, its standard deviation $\hat{s}(\mathbf{x})$, the posterior probability $P[C^j \mid \mathbf{x}]$ for each class, $j = 1, ..., N_c$, and the class indicator $d(\mathbf{x})$. For this purpose, SANN is structured as shown in Fig. 2. It consists of four layers, namely *input layer, GKF layer, summation layer, and estimator layer*, in which the neurons or nodes between layers are interconnected successively by feed-forward direction

In the following paragraphs, the function and

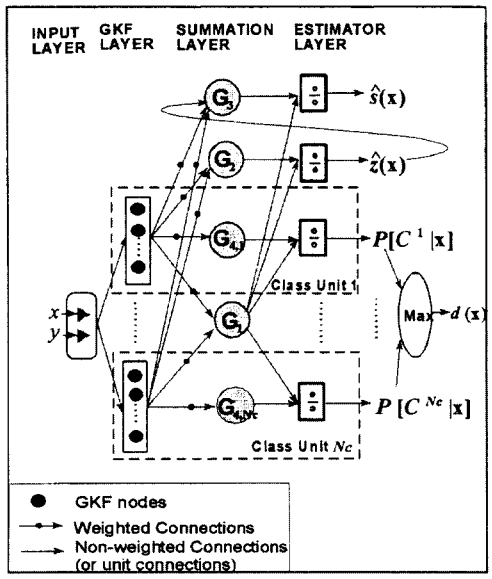

Fig. 2. Structure of the spatial analysis and neural network (SANN) model. GKF stands for Gaussian kernel function.

the connection mechanism of each layer will be explained in detail. Considering a two-dimensional domain, the input layer has two nodes which represent the x and y coordinates, i.e. the vector $\mathbf{x} = [x, y]$. The connections of the input layer implement a pass of the input coordinate vector $\mathbf{x} = [x, y]$ to the *GKF layer*, and those are not weighted. The GKF layer consists of N Gaussian Kernel Function (GKF) nodes. To determine the posterior probability estimator and the Bayesian classifier, the GKF nodes must be divided up into N_c class units as shown in Fig. 2. For doing this, the observed set $\{ \mathbf{X}_n, Z_n \mid n = 1, ..., N\}$ is rearranged as $\{ \mathbf{X}_{(kj)}, Z_{(kj)} \mid k = 1, ..., N^j$ and $j = 1, ..., N_c\}$. $\mathbf{X}_{(kj)}$ is then located at the center of the k-th GKF node in class unit j in which the number of the GKF nodes is N^j. Then, the transfer or activation functions of the k-th GKF node in class unit j are expressed as:

$$a_{(k,j)} = \exp[-\frac{D^2_{x(k,j)}}{2\sigma^2_{x(k,j)}}] \quad (1)$$

where $a_{(j,k)}$ = the GKF node output from the k-th node in class unit j : $D_{x(k,j)}$ = the Euclidean distance between the input vector $\mathbf{x}$ and the k-th center $\mathbf{X}_{(kj)}$ in class unit j and the square of it is expressed as $D^2_{x(k,j)} = (\mathbf{x}-\mathbf{X}_{(k,j)})^T (\mathbf{x}-\mathbf{X}_{(k,j)})$; and $\sigma_{x(k,j)}$ width for the k-th GKF node in class unit j. Each GKF node has the internal parameters; $\mathbf{X}_{(kj)}$ = the position of the center of the GKF node in two-dimensional space, and = the smoothing parameter known as the width of the GKF nodes. The function of the GKF node may be summarized as: the output from each GKF node is a function of the Euclidean distance from the center $\mathbf{X}_{(kj)}$ to the input point $\mathbf{x}$, and each GKF node only responds (or activates) when the input pattern falls within its *receptive field* which is defined by the *width* of the GKF node (Poggio and Girosi, 1990). When the input vector $\mathbf{x}$ is placed at the center of the GKF node $\mathbf{X}_{(kj)}$, the output (1) becomes the maximum value which is one. Otherwise, the magnitude of the GKF output decreases exponentially, as the input vector is farther from the center.

The outputs of the GKF nodes in the GKF layer are passed to the summation layer with the weighted connections. Then, the *summation layer* provides the following outputs:

$$G_1 = \sum_{j=1}^{N_c} \sum_{k=1}^{N^j} \left(\frac{1}{\sigma^2_{x(k,j)}} \right) a_{(k,j)} \quad (2a)$$

$$G_2 = \sum_{j=1}^{N_c} \sum_{k=1}^{N^j} \left(\frac{1}{\sigma^2_{x(k,j)}} \right) Z_{(k,j)}\, a_{(k,j)} \quad (2b)$$

$$G_3 = \sum_{j=1}^{N_c} \sum_{k=1}^{N^j} \left(\frac{1}{\sigma^2_{x(k,j)}} \right) \{ (Z_{(k,j)})^2 - 2Z_{(k,j)}\, \hat{z}(\mathbf{x}) + \hat{z}(\mathbf{x})^2 + \sigma_z^2 \}\, a_{(k,j)} \quad (2c)$$

$$G_{4,j} = \sum_{k=1}^{N^j} \left(\frac{1}{\sigma^2_{x(k,j)}} \right) a_{(k,j)} \quad (2d)$$

where $Z_{(k,j)}$ = observed value corresponding to the k-th GKF node for class unit j, $\hat{z}(\mathbf{x})$ = estimated value at the point $\mathbf{x}$, σ_z = the smoothing parameter or Gaussian kernel width associated with the spatial variable z which must be defined.

The outputs from the summation nodes are passed to the estimator nodes with unit weights. Then, as shown in Fig. 1, the outputs of estimator nodes assign the estimations of the conditional mean $\hat{z}(\mathbf{x})$, its standard deviation $\hat{s}(\mathbf{x})$, and the posterior probability $P[C^j | \mathbf{x}]$ of each class, respectively, by the following activation:

$$\hat{z}(\mathbf{x}) = \frac{G_2}{G_1} \quad (3)$$

$$\hat{s}(\mathbf{x}) = \sqrt{\frac{G_3}{G_1}} \quad (4)$$

$$\hat{P}[C^j | \mathbf{x}] = \frac{G_{4,j}}{G_1} \quad (5)$$

Finally, the class indicator $d(\mathbf{x})$ is determined by assigning the class with maximum posterior probability.

SANN consists of three operation modes, namely, a training mode, an interpolation mode, and a classification mode. In the training mode, the model structure is constructed according to the classes defined by the user as described above. In addition, the model parameters such as the centers and the widths for all GKF

nodes must be determined by using sample observations. The training procedure can be summarized as:

(1) Prepare the observation set $\{\mathbf{X}^n, Z^n \mid n=1, \ldots, N\}$ where N is the number of observations.

(2) Define the classes $C^j = \{C^1, C^2, \ldots, C^{Nc}\}$ with truncation levels $\{TL(j) \mid j=1, \ldots, N_c\}$. Based on the definition of the classes, classify the observation set into each class C^j with $\{\mathbf{X}_{(kj)}, Z_{(kj)} \mid k=1, \ldots, N^j; j=1, \ldots, N_c\}$.

(3) Set the centers of the GKF nodes with the observed coordinate vector $\mathbf{X}_{(k,j)}$. For instance, the center of the k-th GKF node in class unit j is assigned to be $\mathbf{X}_{(k,j)}$. Here, the class layer is arranged with N_c class units as shown in Fig. 2.

(4) Determine the widths $\sigma_{x(k,j)}$ of the GKF nodes. The widths represent the shape of the Gaussian kernel as well as the diameter of the receptive region. They have a profound effect upon the accuracy of the estimation (Haykin, 1994). To cover the whole input space as uniformly as possible, $\sigma_{x(k,j)}$ must be small when the distances between centers are close. Otherwise, it must be large when the centers are separated far away from each other. In this study, the *P-nearest* neighbor method (Moody and Darken, 1989) is applied to determine $\sigma_{x(k,j)}$ where P is the number of the nearest neighbor points. First, the root mean square distance (RMSD) between a center $\sigma_{x(k,j)}$ and its *P-nearest* neighbors is determined for each GKF node:

$$\mathrm{RMSD}_{(k,j)} = \sqrt{\frac{1}{P}\sum_{i=0}^{P} \left| \mathbf{X}_i - \mathbf{X}_{(k,j)} \right|^2} = \sqrt{\frac{1}{P}\sum_{i=0}^{P} (\mathbf{X}_i - \mathbf{X}_{(k,j)})^T (\mathbf{X}_i - \mathbf{X}_{(k,j)})} \quad (6)$$

where $\mathbf{X}_i$ is the i-th nearest neighbor point from the center $\mathbf{X}_{(k,j)}$ of the k-th GKF node in class unit j. Then the width $\sigma_{x(k,j)}$ is given by $\sigma_{x(k,j)} = \mathrm{RMSD}_{(k,j)} / F$ where F is a control factor. Saha and Keeler (1990) stated that just one nearest neighbor, i.e. $P=1$ can produce the desired performance.

(5) After setting the centers and the widths of the GKF nodes, the estimates at the observed points are obtained as $\hat{z}(\mathbf{X}_{(k,j)}) = G_2/G_1$. Then, the root mean square error (RMSE) between the observed values $Z_{(k,j)} = z(\mathbf{X}_{(k,j)})$ and the estimated values $\hat{z}(\mathbf{X}_{(k,j)})$ is determined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N_c}\sum_{k=1}^{N^j} [Z(\mathbf{X}_{(k,j)}) - \hat{z}(\mathbf{X}_{(k,j)})]^2} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} [Z_n - \hat{z}(\mathbf{X}_n)]^2} \quad (7)$$

Then, the width of the spatial variable z, σ_z is determined by $\sigma_z = \mathrm{RMSE}$.

Once the training is completed, the interpolation mode is performed as:

(1) Enter the set of spatial coordinate vectors $\{\mathbf{x}^m \mid m=1, \ldots, M\}$ where m is a given point in the region and M is the number of interpolation points.

(2) Obtain the interpolated value $\hat{z}(\mathbf{x}^m)$, the standard deviation of the estimate $\hat{s}(\mathbf{x}^m)$, the observation point density $\rho(\mathbf{x}^m)$, and the posterior probability $P[C^j \mid \mathbf{x}^m]$ for each class.

After completing the interpolation mode, then the classification mode is accomplished by using the estimated posterior probabilities.

3. Regional Mean Annual Precipitation Analysis

The area selected for this paper is the whole

area of South Korea. It is located between the 126 and 130 degrees East longitude and the 34 and 38.5 degrees North latitude. Mean Annual Precipitation (MAP) data was calculated using the periods of 1966-1996 at 39 precipitation stations as shown in Figure 3. In this figure,

Table 1. Precipitation Stations and Mean Annual Precipitation at South Korea.

Site No.	Station ID	Station Name	(Longitude, Latitude)	Mean Annual Precipitation, MAP(mm)
1	90	Sokcho	(128.6, 38.3)	1291
2	100	Taeganryung	(128.8, 37.7)	1491
3	101	Chunchun	(127.7, 37.9)	1247
4	105	Kangreung	(128.9, 37.8)	1377
5	108	Seoul	(127, 37.6)	1310
6	112	Inchun	(126.6, 37.5)	1134
7	119	Suwon	(127, 37.3)	1250
8	129	Susan	(126.5, 36.7)	1186
9	130	Uljin	(129.4, 37)	1051
10	131	Chungju	(127.4, 36.6)	1200
11	133	Taejun	(127.4, 36.3)	1257
12	135	Chupungryung	(128, 36.2)	1130
13	138	Pohang	(129.4, 36)	1093
14	140	Kunsan	(126.7, 36)	1147
15	143	Taegu	(128.6, 35.9)	1010
16	146	Junju	(127.2, 35.8)	1247
17	152	Ulsan	(129.3, 35.6)	1272
18	156	Kangju	(126.9, 35.2)	1314
19	159	Pusan	(129, 35.1)	1453
20	162	Tongyoung	(128.4, 34.9)	1354
21	165	Mokpo	(126.4, 34.8)	1075
22	168	Yeosu	(127.7, 34.7)	1386
23	192	Jinju	(128.1, 35.2)	1391
24	201	Kanghwa	(126.5, 37.7)	1128
25	211	Inje	(128.2, 38.1)	946
26	212	Hongchun	(127.9, 37.7)	1118
27	214	Samchuk	(129.2, 37.4)	1088
28	221	Jechun	(128.2, 37.2)	1137
29	223	Chungju	(127.9, 37)	1028
30	232	Asan	(127, 36.8)	1064
31	235	Boryung	(126.6, 36.3)	1076
32	245	Jungeup	(120.9, 35.6)	1092
33	256	Sunchun	(127.6, 35.1)	1274
34	272	Youngju	(128.5, 36.9)	1035
35	277	Youngduk	(129.4, 36.5)	893
36	278	Eusung	(128.7, 36.4)	850
37	281	Youngchun	(129, 36)	867
38	284	Kuchang	(127.9, 35.7)	1087
39	288	Milyang	(128.8, 35.5)	1062

Basic Statistics of MAP	
Mean	1165
Standard Deviation	157
Coefficient of Variation	0.13
Coefficient of Skewness	0.06
75 % percentile	1256
50 % percentile	1135
25 % percentile	1079
Minimum	850
Maximum	1491

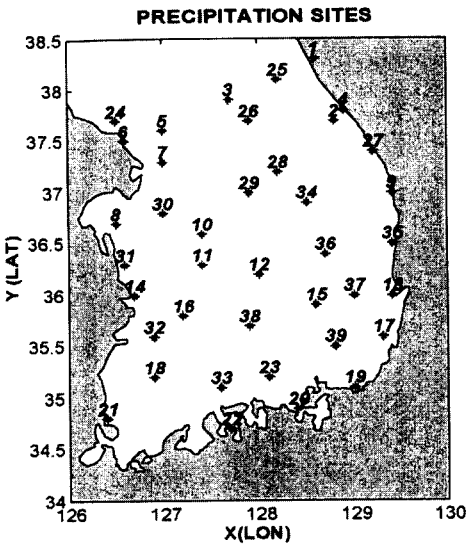

Fig. 3. South Korea and Selected Precipitation Sites

the precipitation sites were denoted by the serial numbers which indicated in Table 1. Table 1 shows the station name, serial number, station number for Korean Meteorological Agency, elevation(m), and MAP data (mm) for each station. As stated before, Thiessen, Inverse Distance, and Kriging methods have been commonly used for spatial analysis of certain hydrologic problems such as precipitation. However, the superiority of SANN against those methods was confirmed in Shin and Salas (1997) based on several real hydrological and environmental applications.

SANN was applied to interpolate MAP over the study region and the standard deviation of the estimate. In addition, the region is classified into wet, normal, and dry areas. At first, the SANN was trained based on the 39 MAP data observed in the study region as follows:

(1) the 39 observed values $Z_{(k,j)}$ and its corresponding coordinate vectors $\mathbf{X}_{(k,j)}$ were classified into three classes C^j, j=1,2,3, using two truncation levels, $TL(1)$ = 1078 mm (25 % percentile) and $TL(2)$ = 1256 mm (75 % percentile) where the percentiles are derived from the 39 values of MAP;

(2) the centers of GKF nodes were made equal to the corresponding coordinate vectors;

(3) the widths of the GKF nodes $\sigma_{x(k,j)}$ were determined using the control parameters P = 1, and F = 1.6;

(4) using the determined centers and widths, the estimates at the 36 observed points were made. Then, the root mean square error (RMSE in Eq. (7)) between the observed and estimated values was 12 mm and it was used for the value of σ_z.

After the training of SANN, the estimates of spatial properties such as MAP field $\hat{z}(\mathbf{x}^m)$, the standard deviation of the estimate $\hat{s}(\mathbf{x}^m)$, the posterior probabilities $P[C^j | \mathbf{x}^m]$ were constructed on the grid system which a cell is a 0.2 degree × 0.2 degree rectangular. Table 2 shows the areal statistical properties for each estimated properties of the MAP field at the whole South Korea.

Comparing the statistics of the observed MAPs and the estimated MAP, $\hat{z}(\mathbf{x}^m)$ as shown in Table 2, the areal average of the estimated MAP field was 1170 mm which is almost close to the value 1165 mm for the observed field. In addition, the standard deviation and the skewness coefficient of the estimated MAP field were closed to the observed field. This indicates that the SANN estimates of the MAP field reproduced well the overall distribution of the observed data. The areal average of standard deviation of the estimates which indicates the areal estimation error was 5 % of the areal average MAP 1170 mm.

Fig. 4 (a) shows the areal distribution of the estimated MAP over the entire region (interpolated map). It indicates that the middle region of Kyung-gi Province around the

Table 2. Summary of the Basic Statistics for Interpolation and Classification Based on SANN for the Mean Annual Precipitation at South Korea

	Observed MAP (mm)	$\hat{z}(\mathbf{x}^m)$ (mm)	$\hat{s}(\mathbf{x}^m)$ (mm)	$P(C^1 \mid \mathbf{x}^m)$	$P(C^2 \mid \mathbf{x}^m)$	$P(C^3 \mid \mathbf{x}^m)$
Mean	1165	1170	64	0.42	0.37	0.21
Standard Deviation	157	149	37	0.40	0.33	0.23
Coef. Of Variation	0.13	0.13	0.59	0.95	0.89	1.10
Coef. Of Skewness	0.06	0.08	1.09	0.63	0.62	0.70

Note: $\hat{z}(\mathbf{x}^m)$ = interpolated MAP at $\mathbf{x}^m$

$\mathbf{x}^m$ = interpolation points in 2 km × 2 km grid system

$\hat{s}(\mathbf{x}^m)$ = standard deviation of the estimates at $\mathbf{x}^m$

$\rho(\mathbf{x}^m)$ = observation point density at $\mathbf{x}^m$

$P(C^j \mid \mathbf{x}^m)$ = posterior probability that the MAP at $\mathbf{x}^m$ is in Dry, Normal, and Wet areas, respectively, j = 1, 2, 3.

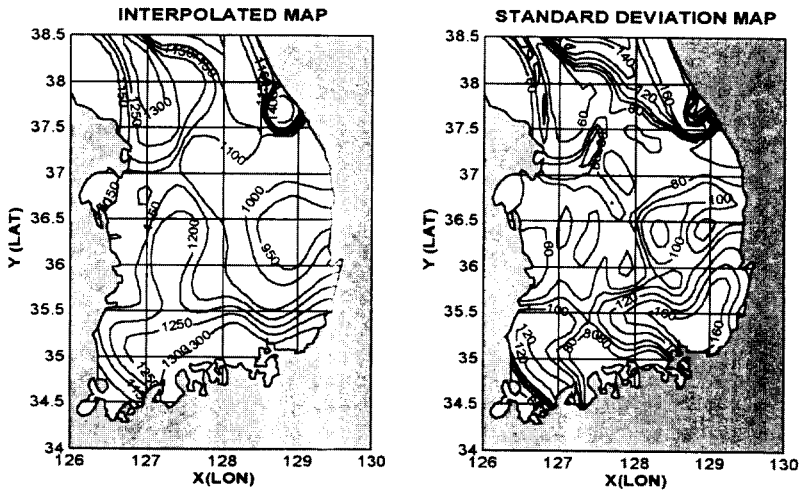

Fig. 4. Interpolated Field (a) and Standard Deviation of Estimate Field (b) for the MAP at South Korea Constructed by SANN Model.

coordinate (127, 38), the east ocean side, and the south ocean side in Kyung-sang-nam Province and Jun-ra-nam Province were the high MAP regions. In the contrast, the Kyung-sang-buk Province around the coordinate (129, 36.5), and the southwestern ocean side in Jun-ra-nam Province around the coordinate (126.5, 34.5) presents the low MAP regions. In terms of the estimation error of MAP field, the areal distribution of standard deviation of the estimates was made as in Fig. 4(b). The result indicates that the high error of estimation about 140 - 160 mm appears at the southeastern region around (129, 35.5) and the northeastern region around (128.5, 38), because these regions are located between the high

MAP and low MAP regions. In addition, the region around (126.8, 35) where the observation MAP sites are scattered comparing with the other regions had also high standard deviation of estimation about 120. Overall, the standard deviation estimator of SANN was performed well for the MAP field.

Based on the posterior probability estimator of SANN, the point dry, normal, and wet probabilities were defined as

Point Dry Probability: the probability that the arbitrary point $\mathbf{x}^m$ will be classified into dry region

$$P[C^1 \mid \mathbf{x}^m] = Pro(\text{MAP} \le 1078\,\text{mm} \mid \mathbf{x}^m) \tag{8a}$$

Point Normal Probability: the probability that the arbitrary point $\mathbf{x}^m$ will be classified into normal region

$$P[C^2 \mid \mathbf{x}^m] = Pro(1078\,\text{mm} < \text{MAP} \le 1256\,\text{mm} \mid \mathbf{x}^m) \tag{8b}$$

Point Wet Probability: the probability that the arbitrary point $\mathbf{x}^m$ will be classified into wet region

$$P[C^3 \mid \mathbf{x}^m] = Pro(\text{MAP} \le 1256\,\text{mm} \mid \mathbf{x}^m) \tag{8c}$$

Here, the first truncation level TL(1) which divides dry and normal classes was MAP = 1078 mm (25 % percentile of MAP) and the second truncation level TL(2) which divides normal and wet classes was MAP = 1256 mm (75 % percentile). The truncation levels may vary based on the individual project and user's purpose.

After estimating each probability over the whole region, the dry probability map, normal probability map, and wet probability map were contoured as in Fig. 5 (a), (b), and (c), respectively. This result may be useful for water engineers for making a probabilistic decision of dry, normal, and wet regions at the arbitrary point. For instance, the probability that the location around the point (128, 35) is considered as a wet region, is over 90 % as shown in Fig. 5 (c) and the location around the point (29, 36) has over 90 % of the probability to be classified into dry region.

Then, based on the Bayesian classifier of SANN model described above, the entire South Korea was classified into dry region, normal

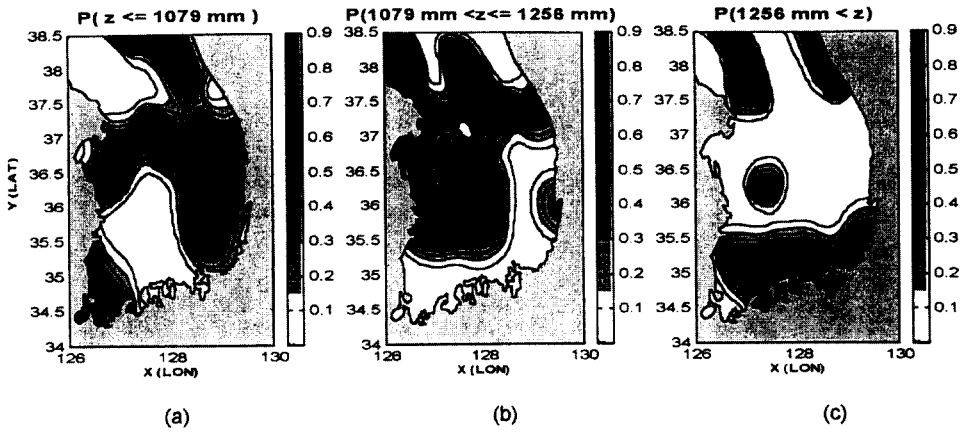

Fig. 5. Posterior Probability Maps for (a) Dry Region, (b) Normal Region, and (c) Wet Region, respectively.

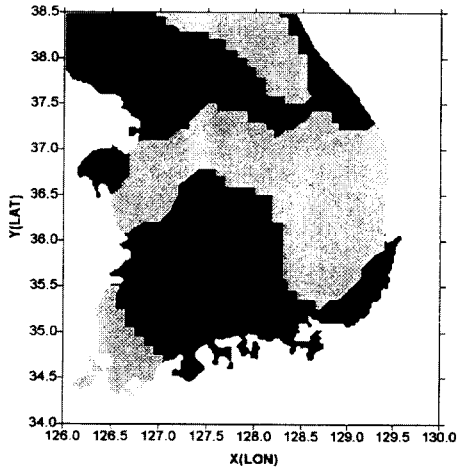

Fig. 6. Classification Map of South Korea to Dry, Normal, and Wet Regions (White: Dry Region, Gray: Normal Region, Black: Wet Region)

region, and wet region as the following criteria:

Dry Area:

$$\text{if } \max P[C^1 \mid \mathbf{x}], P[C^2 \mid \mathbf{x}], \text{ and } P[C^3 \mid \mathbf{x}] = P[C^1 \mid \mathbf{x}] \tag{9a}$$

Normal Area:

$$\text{if } \max P[C^1 \mid \mathbf{x}], P[C^2 \mid \mathbf{x}], \text{ and } P[C^3 \mid \mathbf{x}] = P[C^2 \mid \mathbf{x}] \tag{9b}$$

Wet Area:

$$\text{if } \max P[C^1 \mid \mathbf{x}], P[C^2 \mid \mathbf{x}], \text{ and } P[C^3 \mid \mathbf{x}] = P[C^3 \mid \mathbf{x}] \tag{9c}$$

Fig. 6 shows the Bayesian classification of the whole South Korea into dry (white), normal (gray), and wet (black) regions. As shown in this figure, the middle of Kyung-gi Province around (127,38), and the south oceanic are were indicated as the wet regions. The Kyung-sang-buk Province around (129, 36) and the northern Chung-chung Province appears the dry regions.

4. Summary and Conclusions

In the real hydrological field, the spatial analysis has been of big concern for several aspects such as interpolation of the spatial variable, optimal hydrological network design, and Geographical Information System (GIS) implementation. The primary purpose of this study was to introduce Spatial Analysis Neural Network (SANN) model which is an alternative spatial analysis techniques for an arbitrary spatial variable. Then, the second purpose was to apply the SANN for analyzing the Mean Annual Precipitation (MAP) Field at the South Korean region below 38 latitude degree.

In the problems of water balance and management associated with the spatial variability of MAP, three questions often arise: (1) How can the MAP field is distributed over the region based on the limited observation points? (Interpolation question); (2) What is the probability that an arbitrary point belongs to dry, normal, or wet regions? (probability-of-occurrence question); and (3) What are the boundaries dividing between dry, normal, and wet areas? (optimal-classification problem). Those questions were answered by the interpolator, the posterior probability estimator, and the Bayesian classifier provided by SANN for the South Korean region. In addition, the procedure to be achieved above results was described in detail, which can be used for any region of interest. The results were made for all 0.2 by 0.2 degree points and contoured over the South Korean region, which can help the reader understanding the spatial distribution of the South Korean MAP field, the dry, normal, and wet probabilities, and the classification of the MAP region.

As stated in the main contents, the use of SANN to analyze the spatial MAP field can achieve several advantages such as the nonlinear interpolation based on the nature of neural networks, providing the probability estimation, and optimal spatial classification, which can not be provided by the traditional

techniques such as Thiessen method and Kriging etc.

감사의 글

본 연구는 한국과학재단 및 부산대학교 생산기술 연구소 연구비 지원에 의해 수행되었으며, 본 연구의 2번째 저자는 학술진흥재단의 연구비('95 신진연구인력 연구장려금)을 지원받아 수행하였기에 이에 감사드립니다.

References

ASCE Task Committee on Geostatistical Techniques. (1990). "Review of geostatistics in geohydrology: I. Basis concepts." (1990). *J. of Hydraulic Engineering*, Vol. 116, No. 5, pp. 612-632.

Brass, R.L, and Rodriguez-Iturbe, I. (1985). *Random functions and hydrology*. Addison-Wesley Pub. Comp.

Haykin, S. (1994). *Neural networks: A comprehensive foundation*. Macmillan College Pub. Comp., Inc.

Kassim, A.H.M. and Kottegoda N.K. (1991). "Rainfall network design through comparative Kriging methods." *Hydrological Sciences - Journal des Sciences Hydrologiques*, Vol. 36, No. 3, pp. 431-446.

Markus, M., Shin H.S., and Salas J.D. (1995). "Predicting streamflows based on neural networks." *1995 First International Conference on Water Resources Engineering*, ASCE, San Antonio, TX, 1995.

Moody, J.E., and Darken D.J. (1989). Fast Learning in Networks of Locally-tuned Processing Units, *Neural computation 1*, pp. 281-294.

Novotny, J.V.H., Feng X., and Capodaglio A. (1991). "Time series analysis models of activated sludge plants." *Wat. Sci. Tech.*, Vol. 23, pp. 1107-1116.

Poggio, T., and Girosi F. (1990). "Regularization algorithms for learning that are equivalent to multilayer networks." *Science*, Vol. 247, pp. 978-982.

Rashid, A., Aziz A., and Wong K. F. (1992). "A neural network approach to the determination of aquifer parameters." *Ground Water*, Vol. 30, No. 2, pp. 164-166.

Rogers, L.L. and Dowla F.U. (1994). "Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling." *Water Resour. Res.*, Vol. 30, No. 2, pp. 4578-4581.

Saha, A., and Keeler J.D. (1990). "In algorithms for better representation and faster learning in radial basis function networks." *Advance in Neural Information Processing Systems 2*, Edited by Touretzky, D.S. et al., San Monteo, CA: Morgan Kaufmann, pp. 482-489.

Shin, H.-S. and Salas J.D. (1997). "Spatial analysis neural network model and its applications to hydrological and environmental data." *Water Resources Paper No. 107*, Dept. of Civil Engr., Colorado State University, Ft. Collins, in print.

Tabios, G.Q., and Salas J.D. (1985). "A comparative analysis of techniques for spatial interpolation of precipitation." *Water Resources Bulletin*, Vol. 21, No. 3, pp. 365-380.

(논문번호:98-010/접수:1998.02.16/심사완료:1998.12.02)