신경망을 이용한 우리나라의 시공간적 가뭄의 해석

Spatial-Temporal Drought Analysis of South Korea Based On Neural Networks

신 현 석*
Shin, Hyun Suk

Abstract

A methodology to analyze and quantify regional meteorological drought based on annual precipitation data has been introduced in this paper In this study, based on posterior probability estimator and Bayesian classifier in Spatial Analysis Neural Network (SANN), point drought probabilities categorized as extreme, severe, mild, and non drought events has been defined, and a Bayesian Drought Severity Index (BPSI) has been introduced to classify the region of interest into four drought severities. For example, the proposed methodology has been applied to analyze the regional drought of South Korea. This is a new method to classify and quantify the spatial or regional drought based on neural network pattern recognition technique and the results show that it may be appropriate and valuable to analyze the spatial drought.

요지

본 연구에서는 공간적으로 분포되어 있는 연강우량 자료를 이용한 지역 기상학적인 가뭄을 정의하고 해석하는 모형을 제시한다. 비선형, 비매개변수법에 기초한 공간 해석 신경망 (Spatial Analysis Neural Network: SANN)모형을 이용하여, 각 년에 대하여 공간의 임의 점에서의 극심, 심, 경심, 및 비 가뭄 확률을 전 대상 지역에 대하여 산출을 통하여 가뭄확률도를 작성하며, Bayesian 가뭄 심도 지수 (BDSI)를 통하여 전 대상 지역을 가장 적절하게 극심, 심, 경심, 미 가뭄 지역으로 분류하는 방법을 제시한다. 또한, 각 년의 대표적인 가뭄의 형태를 제시하여줄 수 있는 지역 가뭄 확률과 지역 가뭄 확률 지수를 소개한다. 이 모든 시공간적 가뭄 해석의 방법은 실제로 우리나라 (남한) 전역에 대하여 실시하여, 과거 1967년부터 1996년 까지의 공간적이고 시간적인 가뭄의 발생 현황과 그 특징을 조사한다. 이는 우리나라 장기 수자원 개발 및 유역 관리를 더욱 정량적인 가뭄 정보에 의해 수행하게 하여 줄 수 있을 것이다.

1. Introduction

For the water resource planning and management, defining and analyzing of drought have been of big concern to meteorologist and hydrologist in the several decades. Drought analysis may be made based on single site data (Yevjevich, 1967; Dracup et al, 1980) and multisite data (Tase, 1976, Santos et. al, 1983; Soule, 1992; Guttman et al., 1992) depending on the specific purpose of the study at hand. In this paper, we are concerned with regional drought analysis.

Choosing the variable to define a regional drought depends usually on the purpose of the study. Among various variables such as precipitation, streamflow, soil moisture, and moisture content in the air, precipitation has been commonly applied for meteorological drought analysis (see for instance, Tase, 1976, Santos et. al, 1983, Chang, 1991). Annual precipitation will be considered as the key variable for drought analysis in this study.

Keywords: Drought, Neural Networks, Drought Probability, Bayesian Drought Severity Index, South Korea

* 정회원, 부산대학교 토목공학과 조교수

In general, the historical annual precipitation data estimated in a region arises from data which are observed at a limited number of gaging stations and which are usually of different record lengths. Variability in space may be considered by using an interpolation technique so as to estimate precipitation at any ungaged location in the area. Several interpolation methods such as Thiessen Polygons, Inverse Distance, Multi-quadric, Polynomial, and Kriging have (Salas et al., 1997) been applied to solving water resources problems. Among those, Tase (1976) used the polynomial method to regionalize precipitation data and Chang (1992) used the Kriging method for investigating monthly precipitation droughts. Recently, Shin and Salas (1997a) developed a nonparametric spatial analysis model constructed with a neural network computational scheme, namely Spatial Analysis Neural Network (SANN) and verified for several applications such as groundwater contamination region classification and mean annual precipitation field (Shin, 1997b; Shin, 1988). In that paper, SANN was used for analyzing the spatial variability of annual precipitation.

In literature, various drought severity indices related to precipitation data have been introduced. Gibbs and Maher (1967) developed the concept of deciles of precipitation for drought analysis. The Palmer Drought Severity Index (PDSI) which was developed by Palmer (1965), may be most well known meteorological drought index. The PDSI relates drought severity to the accumulated weighted differences between actual precipitation and the precipitation requirement of evaporation. It has been widely used for a variety of applications (Karl and Quayle, 1981; Guttman et al, 1992; Akinremi et al, 1996). Recently, McKee et al. (1993) introduced the Standardized Precipitation Index (SPI) which is designed to quantify the precipitation deficit for multiple time scales and calculated by taking the difference of the precipitation from the mean for a particular time sclas and then dividing by standard deviation. In this study, based on the posterior probabilities of drought severities at any point in the region such as extreme drought, severe drought, mild drought, and non-drought from SANN, we provide a new drought severity index, namely Bayesian Drought Severity Index (BDSI). The detail description will be appeared below sections.

For defining droughts into various degrees of severity appropriate truncation levels of annual precipitation must be specified. In principle, truncation levels may vary in space and time, but this would require determining annual demands of precipitation varying over the region, data which are not readily available. In addition, the problem becomes too complex because of the variety of geomorphic or climatic conditions over the region and over time. That is why the historical annual precipitation data are usually normalized and standardized to enable the comparison of data on the same basis at various points in space. Thus, we followed the same approach and used constant truncation levels over the region and over time. The South Korea region is used as an example to illustrate the proposed regional drought analysis procedure and its spatial and temporal characteristics of drought is analyzed and quantified.

2. Methodology of Regional Drought Analysis

Consider that annual precipitation data are represented by $R_t(k)$, t=1,..., N, and k = 1,..., S , where N = the length of record, k is a specific site, and S = total number of sites. Note that the actual record length for a given site k may be shorter than N, so N covers the period from the station with the earliest year to the station with the latest year of record. Also site k is located at the point $\mathbf{X}_t(k)$ which is two dimensional coordinate vector. The historical data $R_t(k)$ for each site k are normalized and standardized so to make the data at all stations comparable on the basis of the standard normal distribution. Here, the normalized and standardized annual precipitation data is denoted by $Z_t(k)$ which corresponds to site with coordinate vector $\mathbf{X}_t(k)$. The next step is to estimate the process $z_t(\mathbf{x})$ at any point $\mathbf{x}$ in the area and to estimate a number of point and spatial statistics useful for the problem at hand. For this purpose, we use the Spatial Analysis Neural Network (SANN) model introduced by Shin and Salas (1997).

The drought severity at a point is obtained by partitioning the event $z_t(\mathbf{x})$ into four classes $\{C^1, C^2, C^3, C^4\}$ which are associated with extreme, severe, mild, and non- drought events, respectively. This is accomplished by using three truncation levels $TL(1)$, $TL(2)$, and $TL(3)$. Then, for each year t and for a system of grid points $\mathbf{x}^m$, m = 1,..., M, which are equally spaced over the region, various statistics are obtained such as conditional mean, posterior probabilities $\hat{P}(C^1|\mathbf{x}^m)$, $\hat{P}(C^2|\mathbf{x}^m)$, $\hat{P}(C^3|\mathbf{x}^m)$, and $\hat{P}(C^4|\mathbf{x}^m)$ and a point drought severity indicator $d(\mathbf{x}^m)$. Here, the posterior probabilities are denoted respectively as *Point Extreme Drought Probability* $EDP_t(\mathrm{x}^m)$, *Point Severe Drought Probability* $SDP_t(\mathrm{x}^m)$, *Point Mild Drought Probability* $MDP_t(\mathrm{x}^m)$, and *Point Non-Drought Probability* $NDP_t(\mathrm{x}^m)$. Based on this information, one can obtain drought probability maps for each drought severity by contouring the estimated drought probabilities over the region. Then, each point x^m is assigned a point drought severity $d(x^m)$ by taking the maximum of the referred point drought probabilities. They provide the spatial distribution of drought severity. Here, we denote the point drought severity indicator $d(x^m)$ as a *Bayesian Drought Severity Index* (BPSI)

2.1 Descriptions of Normalization, Standardization, and Truncation Levels

The first step in the analysis is to normalize and standardize the historical annual precipitation data. Many tests are available for testing the null hypothesis of normality (Kottegoda and Rosso, 1997). Here we applied the skewness test of normality. The power- and log-transformations have been widely used to transform the historical annual precipitation data into normal. For instance, Tase (1976) and Kingery (1992) applied the cube-root (p=1/3) transformation to normalize precipitation data. In this study, both the log and the cube-root transformation were applied. Let $Q_t(k)$ be the annual precipitation data after transformation. This data was further standardized as

$$Z_t(k) = \frac{Q_t(k) - \overline{Q}(k)}{S_Q(k)} \tag{1}$$

where $\overline{Q}(k)$ and $S_Q(k)$ are respectively the mean and standard deviation of $Q_t(k)$.

Hence, the normalized and standardized annual precipitation at each site is assumed to be normally distributed with mean zero and standard deviation one. Then constant truncation levels over the region and throughout the historical period are defined to indicate the severity of a drought. Four drought severities are considered, namely C^1, C^2, C^3, and C^4 representing extreme, severe, mild, and non-droughts, respectively. They are defined by the truncation levels determined by $F[TL(1)] = P[Z \leq TL(1)] = 0.15$, $F[TL(2)] = P[Z \leq TL(2)] = 0.30$, and $F[TL(3)] = P[Z \leq TL(3)] = 0.50$, where $F(.)$ represents the standard normal CDF and $TL(.)$ is the quantile for the specified probability, thus the truncation levels obtained from standard normal tables are: $TL(1)$= -1.035, $TL(2)$= -0.385, and $TL(3)$= 0.0

2.2 Spatial Analysis Neural Network (SANN)

A nonparametric spatial analysis model based on a neural network computational scheme has been developed (Shin and Salas, 1997) for point estimation and classification of spatial data. The spatial analysis is based on further developments of Parzen's (1692) nonparametric point density estimators, Bayesian classifier (Bishop, 1995), and the computational scheme based on a multi-layer feed-forward neural network from. For more detail description of SANN, the reader can refer to Shin and Salas (1997a) and the previous paper (Shin, 1998). Note that all notations related with SANN are used in the same manner and the difference for defining the variables are described in detail in this paper.

2.3 Bayesian Drought Severity Index (BDSI)

Considering the four drought severities as described above, the four posterior probabilities estimated at an arbitrary point **x** for year t are denoted as *point extreme drought probability* $EDP_t(\mathbf{x})$, *point severe drought probability* $SDP_t(\mathbf{x})$, *point mild drought probability* $MDP_t(\mathbf{x})$, and *point non-drought probability* $NDP_t(\mathbf{x})$, respectively. They are defined as:

(a) Point Extreme Drought Probability (=the probability of extreme drought given a point **x**)

$$EDP_t(\mathbf{x}) = P[C^1|\mathbf{x}] = P[z_t(\mathbf{x}) \leq TL(1) \mid \mathbf{x}] \tag{2.a}$$

(b) Point Severe Drought Probability (= the probability of severe drought given a point **x**)

$$SDP_t(\mathbf{x}) = P[C^2|\mathbf{x}] = P[TL(1) < z_t(\mathbf{x}) \leq TL(2) \mid \mathbf{x}] \tag{2.b}$$

(c) Point Mild Drought Probability (= the probability of mild drought given a point **x**)

$$MDP_t(\mathbf{x}) = P[C^3|\mathbf{x}] = P[TL(2) < z_t(\mathbf{x}) \leq TL(3) \mid \mathbf{x}] \tag{2.c}$$

(d) Point Non-Drought Probability (= the probability of non-drought given a point **x**)

$$NDP_t(\mathbf{x}) = P[C^4|\mathbf{x}] = P[TL(3) < z_t(\mathbf{x}) \mid \mathbf{x}] \tag{2.d}$$

Table 1. Bayesian Drought Severity at a Point and Bayesian Drought Severity Index (BDSI)

Decision Condition	Bayesian Drought Severity	BDSI
if max ($EDP_t(\mathbf{x})$, $SDP_t(\mathbf{x})$, $MDP_t(\mathbf{x})$, $NDP_t(\mathbf{x})$) = $EDP_t(\mathbf{x})$	Extreme Drought	4
if max ($EDP_t(\mathbf{x})$, $SDP_t(\mathbf{x})$, $MDP_t(\mathbf{x})$, $NDP_t(\mathbf{x})$) = $SDP_t(\mathbf{x})$	Severe Drought	3
if max ($EDP_t(\mathbf{x})$, $SDP_t(\mathbf{x})$, $MDP_t(\mathbf{x})$, $NDP_t(\mathbf{x})$) = $MDP_t(\mathbf{x})$	Mild Drought	2
if max ($EDP_t(\mathbf{x})$, $SDP_t(\mathbf{x})$, $MDP_t(\mathbf{x})$, $NDP_t(\mathbf{x})$) = $NDP_t(\mathbf{x})$	Non Drought	1

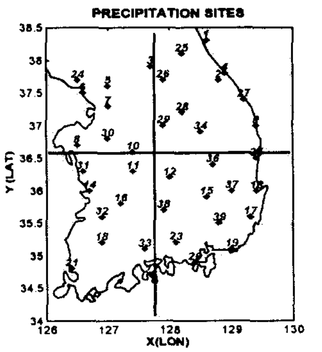
Then, any point $\mathbf{x}$ in the region of interest has a probability of belonging to a certain drought severity, so the Bayesian drought severity at point $\mathbf{x}$ is that which has a maximum probability. Furthermore, the drought severity at a given point is assigned a numerical value $d(\mathbf{x})$ or index as specified in Table 1. Thus Bayesian Drought Severity Indices (BDSI), 3, 2, 1, and 0 represent extreme, severe, mild, and non-drought conditions, respectively. These indices are useful for representing the areal pattern of drought severity at a certain point in time as well as the time evolution of drought severity as a given point.

3. Regional Drought Analysis of South Korea

3.1 Region and Data Description

In this study, the region of South Korea located between 126 and 130 degrees East longitude and 34.0 and 38.0 degrees North latitude was used for regional drought analysis. The total of 39 precipitation stations were assembled with records covering the period 1967 – 1996, which are most recent years. Figure 1 show the South Korean region and the precipitation stations used in this study. As shown in this figure, the whole South Korean region was devided into four representative subregions: East Middle Subregion, West Middle Subregion , Young Subregion, and Honam Subregion.

Fig. 1 Location of the Study Area, South Korea with 39 Precipitation Stations.



The historical annual precipitation data for each station were normalized and standardized following the procedures outlined in Section 2.1. The skewness test of normality was applied to check whether the data was normally distributed. While the data for some of the sites were approximately normal, most of the sites required the cubic root - or log - transformations. After completing the normalization of historical data, the data for each station was standardized based on the equation (1).

3.2 Spatial Distribution of Drought Severity

Based on the Spatial Analysis Neural Network (SANN) model and the procedure to analyze the Baysian Drought Index as described in Section 2.3, the normalized and standardized annual precipitation field at a grid system between 126 and 130 degrees East longitude and 34.0 and 38.0 degrees North latitude were involved in analyzing the spatial distribution of drought severity for each year. Each raster cell is confined in 0.1 degree of latitude and longitude. Each sub-area is assumed to be hydrological and meteorological homogeneous. This assumption implies that every point in a sub-area has approximately the same precipitation characteristics.

For instance, the results for regional drought analysis for the year 1994 are shown in Fig. 2. Figure 2 (a), (b), and (c) show the point drought probability maps for extreme, severe, and mild drought severities, respectively, which were constructed based on the equation (2). These maps display the spatial distribution of drought severity. They may be useful for making a probabilistic statement of drought occurrence and severity at any arbitrary point or area. For instance, Youngnam and Honam subregions have ober 90 % probability that the these subregions were experienced by extreme drought. In addition, Fig. 2 (d) displays the spatial distribution of drought severity based on the Bayesian Drought Severity Index (BDSI) based on Table

1. They are indicated by the color scale with 4 (black) = extreme drought area, 3 (dark gray) = severe drought area, 2 (light gray)= mild drought area, and 1 (white)= non-drought area. It is shown that the West Middle, Youngnam, and Honam subregions had an extreme drought. Overall most of the region in 1994 was affected by an extreme drought. The same graphical results were obtained for all years of the period 1967 - 1996.

Figure 3 shows the results of regional drought analysis for the period of 1967 – 1996 for the whold region of South Korea. These figures may be useful to recognize the spatial distribution of drought severities for each year and help someone to define the extreme, severe, mild, or non drought regions for each year. For instance of the year 1968, the Honam subregion experienced extreme drought and the Youngnam sub region was affected by severe drought, while the East and West subregions experienced mild drought. In addition, the Youngnam and Honam subregions were affected by extreme drought at the year 1995, while the East and West Middle subregions did not experienced the drought. Overall, the years of 1967 – 1968, 1973, 1976-1977, 1982, 1988, 1992, and 1994 – 1996 were affected by severe drought. The last column of Table 3 indicates the subregions where the severe droughts were experienced.

4. Summary and Conclusions

In this paper, we have two purposes: (1) to develop and introduce an approach to analyze and quantify the regional meteorological drought for the region of interest based on annual precipitation data; (2) to apply the proposed method to the entire region of South Korea. After normalization and standardization, the annual precipitation data were classified into four classes (drought severities), namely: extreme drought, severe drought, mild drought, and non-drought. The definition of these classes were done based on three truncation levels corresponding to the 15%, 35%, and 50% quantiles of the standard normal distribution. The posterior probabilities of each drought severity for a given point **x** in the region were determined, i.e. $P(C^j \mid \mathrm{x}), j=1,2,3,4$, and the point was assigned a (point) drought index $d(\mathbf{x})$ either as 4, 3, 2, or 1, depending on whether the maximum posterior probability corresponded to either extreme, severe, mild, or non-drought, respectively. We called the index as "Bayesian Point Drought-Severity Index (BPDI)". This information was useful for constructing a BPDI map and displaying the spatial variability of drought severity for the whole region on a yearly basis.

The proposed regional drought analysis approach was applied to analyze and quantify the regional meteorological droughts for the South Korean region. Annual precipitation data at 39 sites available for the period 1967 – 1996 were utilized. Based on the proposed method, several information related with the regional drought of South Korea was produced: the point drought probability maps and the BDSI map to visualizing the spatial pattern of droughts for each year in the period 1967-1996. Then, the South Korea region was classified into extreme, severe, mild, and non drought severities spatially and yearly. The results obtained above suggest that the proposed approach is valuable and practical tool for defining and analyzing regional droughts.

References

Akinremi, O. O., McGinn, S. M., and Barr, A. G. (1996). " Evaluation of the Palmer Drought Index on the Canadian Prairies," *Journal of Climate*, Vol. 9, 897-905.

Chang, T. J. (1991). "Investigation of Precipitation Drought by Use of Kriging Method," *J. of Irrigation and Drainage Engineering*, 117(6).

Clausen, B. and Pearson, C. P. (1995). "Regional Frequency Analysis of Annual Maximum Streamflow Drought," *Journal of Hydrology*, 173, 111-130.

Dracup, J. A., K. S. Lee, and E. G. Paulson Jr. (1980). " On the Statistical Characteristics of Drought Events," *Water Resour. Res.*, 16(2), 289-296.

Gibbs, W. J. and Maher, J. V. (1967). " Rainfall Deciles as drought Indicators," *Bureau of Meteorology Bulletin No. 48*, Melbourne, Australia.

Guttman, N. B., Wallis, J. R., and Hosking, J. R. M. (1992). "Spatial Comparability of the Palmer Drought Severity Index," Water Resources Bulletin, 28(6), 1111 – 1119.

Karl, T. R. and Quayle, R. G. (1981). " The 1980 summer Heat Wave and Drought in Historical Perspective," *Monthly Weather Review*, 109(10), 2055-2073.

Kingery Jr., R. K.(1992). *A Stochastic Analysis of Spatial Droughts in Colorado*, M.S. thesis, Colo. State

University, Fort Collins, Co.
MacKee, T. B., Doesken, N. J., and Kleist, J. (1993). " The Relationship of Drought Frequency and Duration to Time Scales, *8 th Con. on Applied Climatology*, Anaheim, CA.
Musavi, M. T., K. H. Chan, D. M. Hummels, K. Kalantri, and W. Ahmed. (1992). " A Probabilistic Model for Evaluation of Neural Networks Classifiers," *Pattern Recognition*, 25(10), 1241-1251.
Palmer, W. C. (1965). " Meteorological Drought," *Research Paper No. 45, U.S. Weather Bureau*.Shin, H.-S. and Salas, J. D.(1997a). "Spatial Analysis and Neural Network Modeling for Hydrological and Environmental Data," *Water Resources Research*, sumitted.
Shin, H.-S. (1997b). "Use of the Spatial Analysis Neural Network (SANN) Method for Regional Groundwater Contamination Decision-Making", *J. of Environmental Hydrology*, 5(4).
Soule, P. T.(1992). " Spatial Patterns of Drought Frequency and Duration in the Cintigous USA Based on Multiple Drought Event Definitions," *International Journal of Climatology*, Vol. 12, 11-24.
Tase, N.(1976). *Area-Deficit-Intensity Chracteristics of Drought*, Ph.D. dissertation, Colorado State University, Fort Collins, CO.
Yevjevich, V. (1967), "An Objective Approach to Definitions and Investigations of Conditional Hydrologic Droughts," *Hydrology Paper 23*, Colorado State University, Fort Collins, CO.

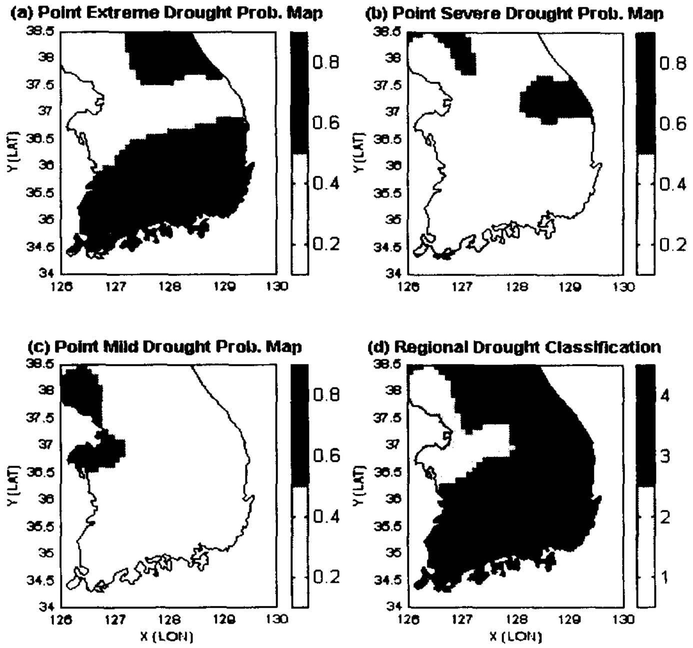


Fig. 2 Regional Drought Analysis for Year 1994

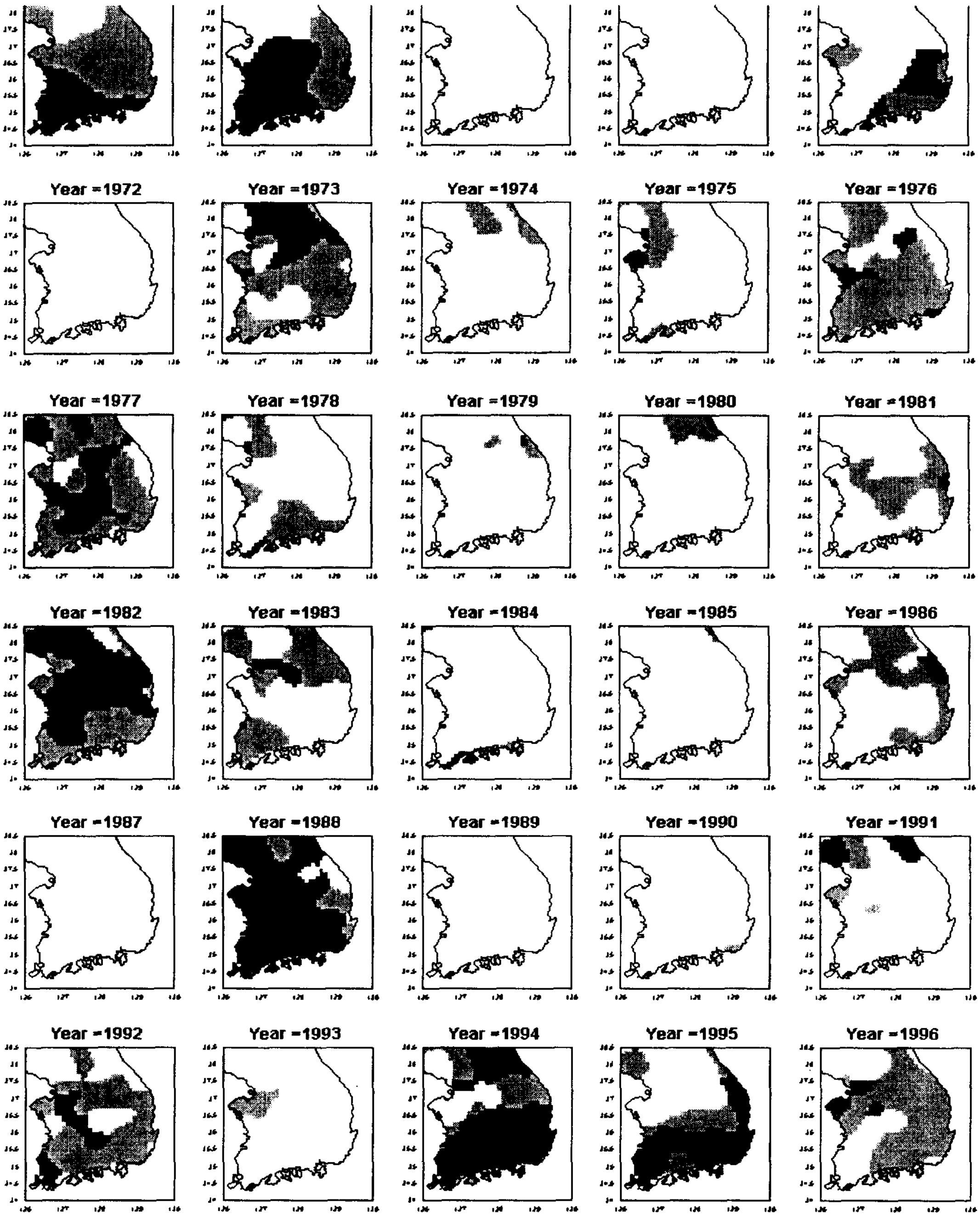


Fig.3 Regional Drought Classification for the Period of 1967 – 1996 at South Korea (White: Non Drought Region, Light Gray: Mild Drought Dark Gray: Severe Drought, Black: Extreme Drought region)