HYDROLOGICAL PROCESSES
Hydrol. Process. **19**, 937–946 (2005)
Published online 21 October 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.5550

Development of regional design hyetographs

Gwo-Fong Lin,* Lu-Hsien Chen and Shih-Chieh Kao
Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan

Abstract:

A method is proposed to establish regional design hyetographs for facilitating the determination of design hyetographs at ungauged sites. The method is applied to the central area of Taiwan. First, the single-station design hyetographs at all rain gauges are analysed using principal components analysis and cluster analysis. The principal components analysis shows that there are six dominant factors, and the cluster analysis indicates that the time to peak rainfall has the largest influence on the classification of hyetographs. It also shows that the single-station hyetographs in the study area can be classified into three clusters. Finally, the homogeneous regions for these three clusters are delineated and the corresponding regional design hyetographs are proposed. Once the homogeneous regions and the regional hyetographs are available, the design hyetograph at the point of interest can be easily determined. The proposed method is expected to be useful for providing the design hyetographs at ungauged sites. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS design hyetograph; principal component analysis; cluster analysis; homogeneous region

INTRODUCTION

A design hyetograph is the time distribution of rainfall during a storm. Once the design point rainfall with certain duration and return period is obtained, the total depth is distributed over the specific duration. The design hyetograph serves as input to a hydrologic system from which the runoff hydrograph is yielded. The choice of the design hyetograph will have a significant influence on the shape and peak value of the hydrograph. Hence, the determination of design hyetographs is an important task in the hydrologic designs. For the establishment of design hyetographs, several methods are available in the literature. According to Chow *et al*. (1988), the methods can be classified into two categories. In the first category the design hyetographs are produced from storm events, whereas in the second category they are developed from intensity–duration–frequency curves. Representative methods in the first category include those proposed by Huff (1967), Pilgrim and Cordery (1975) and Yen and Chow (1980). The Chicago method (Keifer and Chu, 1957) and the alternating block method (Chow *et al*., 1988) are representatives of the second category. In addition to the traditional design hyetographs, there are a number of alternatives. For example, Koutsoyiannis and Foufoula-Georgiou (1993), Garcia-Guzman and Aranda-Oliver (1993) and Cheng *et al*. (2001a) developed design hyetographs using a stochastic approach.

The aforementioned methods are only applicable to gauged sites. To obtain a design hyetograph at an ungauged site, the hyetographs of the nearby rain gauges are often employed. However, is the choice of the hyetographs objective enough? If there is no gauge or the gauge is far away, how to get the reliable design hyetograph at the point of interest is worthy of investigation. Hence, it is justified to establish regional design hyetographs for facilitating the determination of design hyetographs at ungauged sites.

The objective of this paper is to find a method for the establishment of regional design hyetographs. First, principal components analysis (PCA) is performed to transform the original variables of the design

* Correspondence to: Gwo-Fong Lin, Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan.
E-mail: gflin@ntu.edu.tw

Received 16 June 2003
Accepted 8 January 2004

hyetograph into new variables that are independent and orthogonal. Then, cluster analysis is applied to group the rain gauges into specific clusters. Finally, the homogeneous regions for these clusters are delineated and the corresponding regional design hyetographs are proposed.

THE STUDY AREA

Figure 1 shows the study area that is locally referred to as the central region of Taiwan. The study area has an area of about 10 000 km^2. There are 53 rain gauges available (Figure 1). The average annual precipitation is about 2080 mm.

The rainfall data are collected from the Water Resources Agency's computer archives, in which only storms with rainfall depth over 100 mm day^{-1} or rainfall intensity over 20 mm h^{-1} are stored. The design hyetograph of each rain gauge, which is called the single-station design hyetograph herein, is obtained using the method given in the local technical manual (Cheng *et al*., 2001b). One can also refer to Cheng *et al*. (2001a) for a detailed description of the method. In total there are 53 single-station design hyetographs from which regional hyetographs will be developed. The abscissa of the single-station design hyetograph is dimensionless time with 24 intervals, and the ordinate is the percentage of total rainfall.

PRINCIPAL COMPONENT ANALYSIS

In this section, PCA on 53 single-station design hyetographs is performed. PCA is a linear transformation technique that provides a smaller set of uncorrelated variables (called components) from a set of correlated variables while maintaining most of the information in the original data set. PCA is often used as a preprocessing step to clustering (Everitt, 1993), and it is in an attempt to reduce the number of variables. This is important because it helps to reduce future data-collection costs. Usually, most of the variation in a

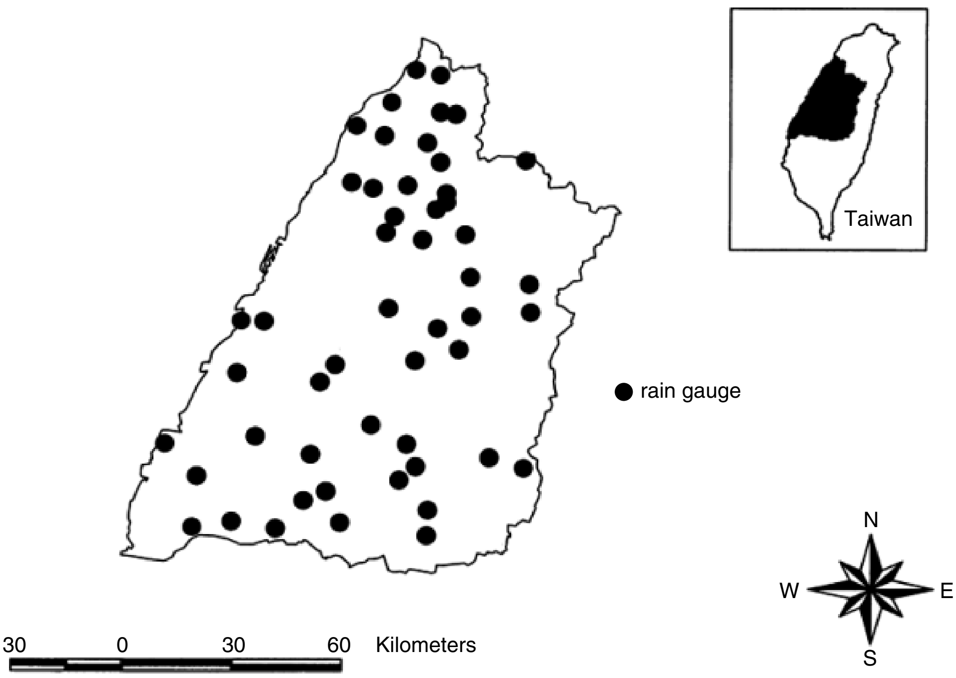

Figure 1. Location of rain gauges in central Taiwan

large group of variables can be captured with only a few principal components. The principles of PCA are described below.

Consider the $p \times n$ data matrix of observations:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \tag{1}$$

where x_{ij} is the jth observation for the ith variable. The sample mean vector is written as

$$\mu_X^{\mathrm{T}} = [\, E(X_1) \quad E(X_2) \quad \cdots \quad E(X_p)\,] \tag{2}$$

and the sample covariance matrix is defined as

$$\Sigma = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \mathrm{Cov}(X_p, X_2) & \cdots & \mathrm{Var}(X_p) \end{bmatrix} \tag{3}$$

where Var( ) and Cov( ) denote the variance and covariance respectively. One can write Equation (3) as

$$\Sigma = C^{\mathrm{T}} \Lambda C \tag{4}$$

where Λ is the diagonal matrix of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} \tag{5}$$

and C is the correlation coefficient matrix between the original variables and the principal components:

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} = [\, C_1 \quad C_2 \quad \cdots \quad C_p\,] \tag{6}$$

Matrix C is orthogonal and thus $C \cdot C^{\mathrm{T}} = I$. The principal components can be written as

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{p1} & y_{p2} & \cdots & y_{pn} \end{bmatrix} \tag{7}$$

The principal components can be presented as linear combinations of the original variables:

$$\begin{cases} y_{1j} = c_{11}x_{1j} + c_{12}x_{2j} + \cdots + c_{1p}x_{pj} \\ y_{2j} = c_{21}x_{1j} + c_{22}x_{2j} + \cdots + c_{2p}x_{pj} \\ \qquad \vdots \\ y_{pj} = c_{p1}x_{1j} + c_{p2}x_{2j} + \cdots + c_{pp}x_{pj} \end{cases} \tag{8}$$

In matrix notation, Equation (8) becomes

$$Y = C \cdot X \tag{9}$$

In Equation (8), the linear coefficients are called the component loadings, i.e. the correlation coefficients between the original variables and the principal components. According to the component loading, the degree of relationship between transformed and original data can be obtained and explained. Furthermore, the covariance matrix of Y becomes

$$\text{Cov}(Y) = C\Sigma C^{\text{T}} \tag{10}$$

Inserting Equation (4) into Equation (10) gives

$$\text{Cov}(Y) = CC^{\text{T}}\Lambda CC^{\text{T}} = \Lambda \tag{11}$$

According to Equation (11), one can find that each principal component is independent and orthogonal to all the others. This means that each principal component presents completely new information.

In this paper, a single-station design hyetograph for a certain rain gauge can be regarded as an observation with 24 variables. The principal components are ordered in such a way that the variance explained by the first principal component is the greatest, the variance explained by the second one is smaller, and so on, and that of the last one is the smallest. The principal components represent the feature pattern of hyetographs. Hence, designable feature patterns (including peak rainfall) of hyetographs can be discriminated when we group rain gauges into clusters. The results of the PCA are presented in Table I, which shows that the first six principal components explain over 90% of the information. That is, one can use only the six principal components to describe all the single-station design hyetographs.

Table I. Results of PCA on 53 single-station design hyetographs

Principal components	Eigenvalue	Variance explained (%)	Total variance explained (%)
1	45·91	51·56	51·56
2	17·23	19·35	70·91
3	6·75	7·58	78·49
4	5·06	5·68	84·17
5	4·09	4·59	88·76
6	2·70	3·03	91·79
7	2·05	2·30	94·09
8	1·52	1·71	95·80
9	0·99	1·10	96·90
10	0·82	0·92	97·82
11	0·44	0·49	98·31
12	0·36	0·40	98·71
13	0·25	0·28	98·99
14	0·16	0·18	99·17
15	0·16	0·17	99·34
16	0·15	0·17	99·51
17	0·11	0·12	99·63
18	0·09	0·10	99·73
19	0·08	0·09	99·82
20	0·05	0·06	99·88
21	0·05	0·05	99·93
22	0·04	0·05	99·98
23	0·02	0·02	100·00
24	0·00	0·00	100·00

CLUSTER ANALYSIS

After principal components analysis is performed, the data sets of principal components of rain gauges are used in cluster analysis to group rain gauges into clusters. Cluster analysis is the determination of natural groupings of similar objects within a multivariate data set. This analysis reduces a large and complex data set to a small number of data groups where members of a group share similar characteristics. The data usually have different scales and importance, and they should be normalized to avoid different weights of data. The normalization formula is

$$z_{ij} = \frac{y_{ij} - \overline{y}_i}{s_i} \qquad i = 1, 2, \ldots, p; j = 1, 2, \ldots, n \tag{12}$$

where i is an index corresponding to the principal component, j is an index corresponding to the data set, p is the number of principal components, n is the number of data sets (i.e. number of total rain gauges in this paper), y_{ij} is the jth transformed observation of the ith principal component, and $\overline{y}_i$ and s_i are respectively the mean and the standard deviation of the ith principal component.

The similarity must be measured, which can be presented using the distance. The distance d_{jk} between the jth and the kth principal components is

$$d_{jk} = \left[\sum_{i=1}^{p}(z_{ij} - z_{ik})^2\right]^{1/2} \tag{13}$$

The distance is constructed in the orthogonal space and the scales of the principal components are the same.

After the similarity has been measured, the cluster analysis can be used to determine the groupings of similar objects. Nonhierarchical clustering methods are frequently chosen to assign data observations to clusters. The best known of the nonhierarchical clustering methods is the so-called K-means method, which is described below.

The centres of the clusters are initialized by randomly selecting from the data set. Then the data set is clustered in the process of assigning each point to the nearest centre. When the data set has been assigned, the average position of the data points within each cluster is calculated and the cluster centre is then moved to the average position. This process of assigning and averaging is repeated until all the cluster centres no longer move. The process is then said to have converged.

In this paper, the K-means method is used for clustering. We have tested various numbers of clusters, e.g. from two to five, etc. After training the number of clusters and considering the practicability of regional hyetographs, we group the 53 rain gauges into three clusters. Figure 2 and Table II show the results of cluster analysis on 53 single-station design hyetographs. The time to peak rainfall for each rain gauge's design hyetograph is also presented in Table II. The times to peak rainfall for these three clusters are 9–11, 12–13 and 14–16 h. The results show that the time to peak has a great influence on the classification of hyetographs.

DELINEATION OF HOMOGENEOUS REGIONS

As stated in the previous section, the 53 rain gauges are grouped into three clusters. In this section, the area for each cluster will be delineated. Such an area is referred to as a homogeneous region herein. Because the rainfall characteristics at ungauged sites are unknown, it is logical to assume that the correlation of rainfall characteristics between two locations increases with decreasing distance. The delineation procedures are summarized as follows.

If the whole region can be grouped into n clusters, then any spatial point must belong to a certain cluster. We assume that there are m clustering points in the whole region, namely $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$. The probability of point (x, y) is grouped into the ath cluster $(1 \leq a \leq n)$ and is denoted as $P_a(x, y)$. If $P_a(x, y)$ is the largest among $P_1(x, y), P_2(x, y), \cdots, P_n(x, y)$, then point (x, y) is grouped into the ath cluster.

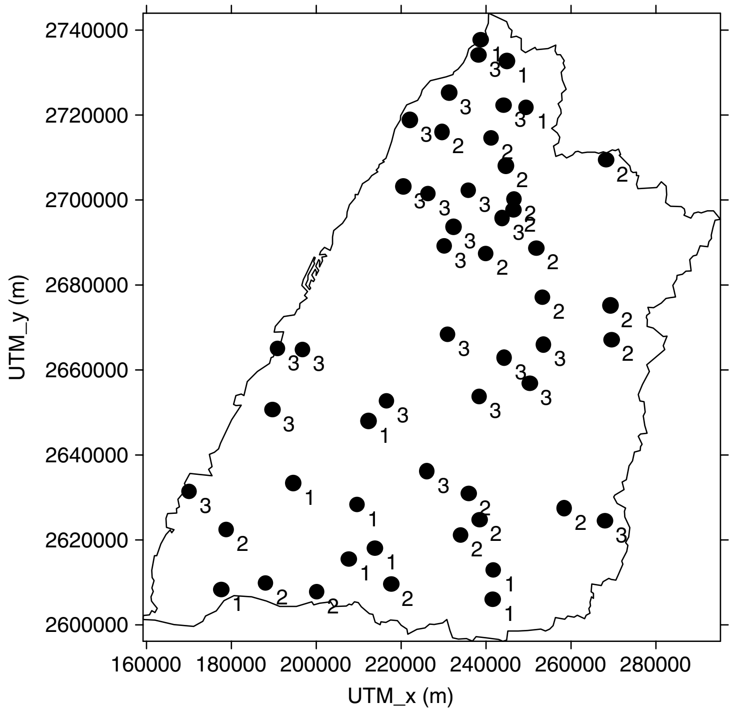

Figure 2. Results of cluster analysis on 53 single-station design hyetographs

To calculate $P_a(x, y)$, first an indicator variable f_h is define as

$$f_h = \begin{cases} 1 & \text{if} \quad h \in a \\ 0 & \text{if} \quad h \notin a \end{cases} \qquad h = 1, 2, \ldots, m \tag{14}$$

The distance between points (x, y) and (x_h, y_h) is

$$d_h = \sqrt{(x - x_h)^2 + (y - y_h)^2} \tag{15}$$

In a like manner, one can obtain the distances $d_1, d_2, \cdots, d_m$. The weight w_h is defined as

$$w_h = \frac{{d_h}^{-2}}{\sum_{h=1}^{m} {d_h}^{-2}} \tag{16}$$

Then the probability that the point (x, y) is grouped into the ath cluster can be written as

$$P_a(x, y) = \sum_{h=1}^{m} w_h f_h \tag{17}$$

In this paper, the study area is divided into a grid pattern. The cluster to which each grid belongs is then determined using Equation (17). That is, the homogeneous region corresponding to a cluster can be delineated. The homogeneous regions for design hyetographs in central Taiwan are shown in Figure 3.

Table II. Results of cluster analysis on 53 single-station design hyetographs

Cluster	UTM_x (m)	UTM_y (m)	Time to peak (h)
1	238 874	2 737 548	11
1	249 184	2 721 730	10
1	244 997	2 732 622	11
1	212 393	2 647 852	10
1	241 748	2 605 846	11
1	241 894	2 612 828	11
1	213 987	2 618 008	9
1	194 720	2 633 204	10
1	209 676	2 628 162	10
1	207 824	2 615 380	11
1	177 766	2 608 246	11
2	244 932	2 708 563	12
2	241 360	2 714 072	13
2	229 739	2 715 960	12
2	268 384	2 709 312	12
2	252 002	2 688 627	13
2	246 676	2 700 040	12
2	246 619	2 697 487	12
2	240 103	2 687 338	13
2	253 414	2 676 967	13
2	269 445	2 675 042	13
2	269 682	2 667 074	13
2	258 551	2 627 348	13
2	234 108	2 620 987	12
2	238 558	2 624 705	14
2	236 015	2 630 736	12
2	217 765	2 609 416	13
2	178 908	2 622 308	13
2	188 186	2 609 772	13
2	200 248	2 607 696	13
3	238 393	2 733 979	13
3	245 021	2 722 253	14
3	235 944	2 702 170	14
3	231 440	2 725 188	13
3	226 507	2 701 352	15
3	222 175	2 718 681	13
3	220 651	2 703 055	14
3	243 913	2 695 642	14
3	232 441	2 693 590	16
3	230 208	2 689 040	14
3	253 414	2 676 967	15
3	250 396	2 656 755	15
3	244 435	2 662 724	15
3	253 671	2 665 892	14
3	238 412	2 653 653	14
3	216 645	2 652 610	14
3	231 026	2 668 212	14
3	196 875	2 664 699	16
3	190 916	2 664 876	14
3	189 782	2 650 483	14
3	268 154	2 624 312	14
3	226 229	2 636 040	14
3	170 168	2 631 257	15

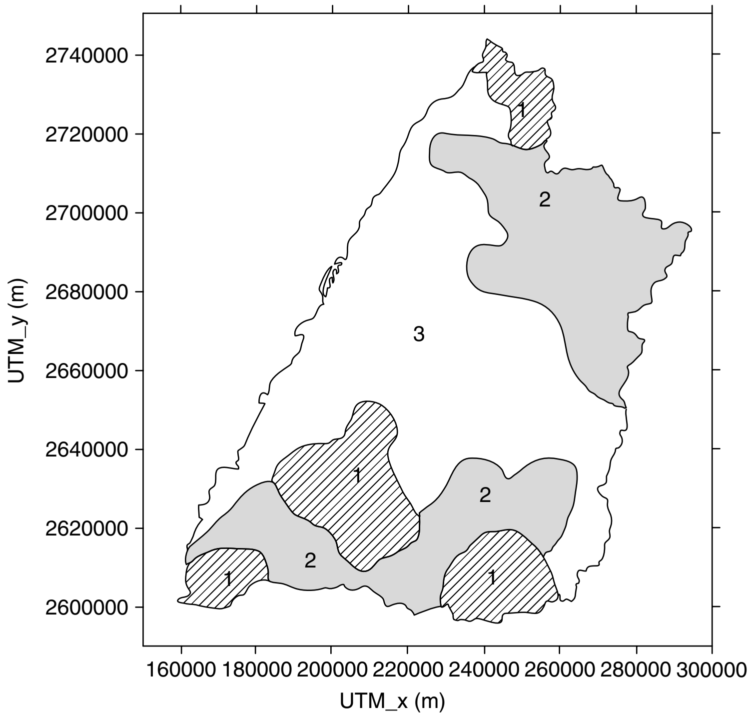

Figure 3. Homogeneous regions of design hyetographs in central Taiwan

REGIONAL DESIGN HYETOGRAPH

Once the homogeneous regions are delineated, the corresponding regional design hyetographs can then be established. If there are N rain gauges in a cluster, then the mean principal components in this cluster can be determined as

$$\bar{z}_i = \frac{1}{N}\sum_{j=1}^{N} z_{ij} \qquad i = 1, 2, \ldots, p \tag{18}$$

where $\bar{z}_i$ is the mean value of the ith principal component, p is the number of principal components and N is the number of data sets in a cluster (i.e. number of rain gauges). Furthermore, let Z^* and $\widetilde{X}$ denote the known matrix of principal components and unknown regional hyetograph respectively. According to Equation (9), $\widetilde{X}$ can be obtained as

$$\widetilde{X} = C^{-1}Z^* \tag{19}$$

The three dimensionless design hyetographs for central Taiwan are shown in Figure 4. The times to peak rainfall of these three regional design hyetographs are 11, 13 and 14 h.

The regional design hyetograph of a homogeneous region is evaluated using two criteria: (1) error of time to peak (ETP) and (2) error of weighting time (EWT).

$$\mathrm{ETP} = \frac{1}{N}\sum_{r=1}^{N}\left(\frac{|T_r - T|}{24}\right) \tag{20}$$

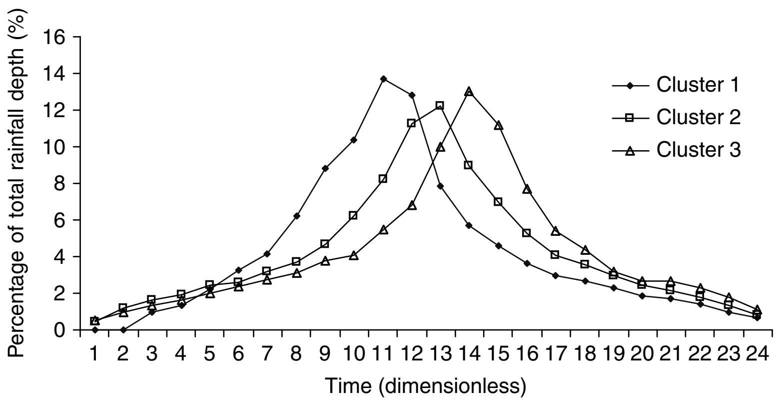

Figure 4. Regional design hyetographs for central Taiwan

where T_r is the time to peak of the rth single-station design hyetograph, T is the time to peak of the regional design hyetograph and N is the number of rain gauges in a certain homogeneous region.

$$\text{EWT} = \frac{1}{N}\sum_{r=1}^{N}\left(\frac{|t_r - t|}{t}\right) \tag{21}$$

where t_r is the time to the centre of mass of the rth single-station design hyetograph and T is the time to the centre of mass of the regional design hyetograph.

The performance of the regional design hyetographs developed is summarized in Table III, wherein the values of ETP and EWT are seen to be very small. Hence, the regional hyetographs can successfully represent the hyetographs for the rain gauges in the same homogeneous region.

SUMMARY AND CONCLUSIONS

A method is proposed for the establishment of regional design hyetographs. First, the single-station design hyetographs for the 53 rain gauges in central Taiwan are established. Then PCA is applied to obtain the principal components of the single-station design hyetographs. It is found that the first six principal components explain over 90% of the information. Based on the transformed data resulting from PCA, cluster analysis (K-means method) is used to group the rain gauges into specific clusters. The 53 rain gauges are grouped into three clusters after training the number of clusters and considering the practicability of regional hyetographs. The results of cluster analysis show that the peak time is the principal factor affecting the classification of hyetographs. Moreover, the regions for three clusters are delineated. Finally, the regional design hyetograph for each cluster (i.e. each homogeneous region) is established. It is found that the regional hyetographs can

Table III. Performance of regional design hyetographs for three regions

Region	Performance	
	ETP (%)	EWT (%)
1	3·70	6·30
2	2·78	3·72
3	4·09	5·96

successfully represent the hyetographs for the rain gauges in the same homogeneous region. The proposed regional hyetographs are very convenient to use. Once the homogeneous region that the point of interest belongs to is determined from Figure 3, the corresponding hyetograph can then be found from Figure 4. The proposed method is expected to be useful for providing the design hyetographs at ungauged sites.

REFERENCES

Cheng KS, Hueter I, Hsu EC, Yen HC. 2001a. A scale-invariant Gauss–Markov model for design storm hyetographs. *Journal of the American Water Resources Association* **37**(3): 723–735.

Cheng KS, Lin GF, Wang RY, Hsu MH, Yu GH, Yu PS, Lee KT. 2001b. *Handbook for hydrological design*. Technical Report, Water Resources Agency, Ministry of Economic Affairs, Taipei, Taiwan (in Chinese).

Chow VT, Maidment DR, Mays LW. 1988. *Applied Hydrology*. McGraw-Hill: New York.

Everitt BS. 1993. *Cluster Analysis*. John Wiley: New York.

Garcia-Guzman A, Aranda-Oliver E. 1993. A stochastic model of dimensionless hyetograph. *Water Resources Research* **29**(7): 2363–2370.

Huff FA. 1967. Time distribution of rainfall in heavy storms. *Water Resources Research* **3**(4): 1007–1019.

Keifer CJ, Chu HH. 1957. Synthetic storm pattern for drainage design. *Journal of the Hydraulics Division, ASCE* **83**(4): 1–25.

Koutsoyiannis D, Foufoula-Georgiou E. 1993. A scaling model of a storm hyetograph. *Water Resources Research* **29**(7): 2345–2361.

Pilgrim DH, Cordery I. 1975. Rainfall temporal patterns for design floods. *Journal of the Hydraulics Division, ASCE* **101**(1): 81–95.

Yen BC, Chow VT. 1980. Design hyetographs for small drainage structures. *Journal of the Hydraulics Division, ASCE* **106**(6): 1055–1076.

