다변량 통계분석을 이용한 준분포형 유출모형 매개변수 지역화

Parameter Regionalization of Semi-Distributed Runoff Model Using
Multivariate Statistical Analysis

이 병 주* / 정 일 원** / 배 덕 호***

Lee, Byong Ju / Jung, Il Won / Bae, Deg Hyo

Abstract

The objective of this study is to suggest parameter regionalization scheme which is integrated two multivariate statistical methods: principal components analysis(PCA) and hierarchical cluster analysis(HCA). This technique is to apply semi-distributed rainfall-runoff model on ungauged catchments. 7 catchment characteristics (area, mean altitude, mean slope, ratio of forest, water content at saturation, field capacity and wilting point) are estimated for 109 mid-sized sub-basins. The first two components from PCA results account for 82.11% of the total variance in the dataset. Component 1 is related to the location of the catchments relevant to the altitude and Component 2 is connected with the area of these. 103 ungauged catchments are clustered using HCA as the following 6 groups: Goesan 23, Andong 6, Imha 5, Hapcheon 21, Yongdam 4, Seomjin 44. SWAT model is used to simulate runoff and the parameters of the model on the 6 gauged basins are estimated. The model parameters were regionalized for Soyang, Chungju and Daecheong dam basins which are assumed as ungauged ones. The model efficiency coefficients of the simulated inflows for these three dams were at least 0.8. These results also mean that goodness of fit is high to the observed inflows. This research will contribute to estimate and analyze hydrologic components on the ungauged catchments.

keywords : Parameter Regionalization, Principal Component Analysis, Hierarchical Cluster analysis, SWAT

요 지

본 연구에서는 미계측유역에 대한 준분포형 강우-유출모형을 적용하기 위한 방법으로 두 개의 다변량 통계기법인 주성분분석과 계층적 군집분석을 연계한 매개변수 지역화 기법을 제안하였다. 109개 중권역 유역에 대해 7개 유역특성인자(유역면적, 평균표고, 평균경사, 산림면적비, 포화토양수분량, 포장용수량, 영구위조점)를 추출하였으며 주성분분석을 수행한 결과 제1, 2 성분이 전체자료의 82.11%를 설명하는 것으로 나타났다. 제1성분은 유역위치, 제2성분은 유역규모와 관계가 있는 것으로 분석되었으며 이들 성분점수로부터 군집분석을 이용하여 103개 미계측유역을 6개 계측유역으로 분류한 결과 괴산댐 23개, 안동댐 6개, 임하댐 5개, 합천댐 21개, 용담댐 4개, 섬진강댐 44개의 미계측유역을 포함하는 것으로 나타났다. 유출모형은 SWAT 모형을 선정하였으며 6개 계측유역에 대한 매개변수를 추정하였다. 매개변수 지역화 결과의 적용성을 평가하기 위해 미계측유역으로 가정한 소양, 충주, 대청댐 상류유역에 대

* 교신저자, 세종대학교 토목환경공학과 박사과정

Corresponding Author, Ph.D. Candidate, Dept. of Civil and Environmental Engrg., Sejong Univ, Seoul 143-747, Korea (e-mail: bjlee0704@paran.com)

** 세종대학교 토목환경공학과 BK21 박사후과정

Post Doctor. Dept. of Civil and Environmental Engrg., Sejong Univ, Seoul 143-747, Korea

*** 세종대학교 물자원연구소 · 토목환경공학과 교수

Prof., Dept. of Civil and Environmental Engrg., Sejong Univ, Seoul 143-747, Korea

해 지역화된 매개변수를 이용하여 유출해석을 수행한 결과 모형효율성계수가 0.8 이상으로 관측치와 적합도가 매우 높게 나타났다. 이상의 결과로부터 다변량 통계분석을 이용한 유출매개변수 지역화 방법은 미계측유역의 유출모의시 활용 가능성을 확인하였다.

핵심용어 : 매개변수 지역화, 주성분분석, 계층적 군집분석, SWAT

1. 서 론

수문분야의 선진국들은 수문현황을 정확하고 효율적으로 파악하기 위해 체계적인 수문관측시스템을 구축하고 있다. 우리나라 또한 수문관측의 중요성을 인지하고 효과적인 수문관측 네트워크를 구축하기 위해 많은 노력을 하고 있다. 우리나라에 설치된 우량/수위 관측소 현황을 살펴보면, 2006년을 기준으로 우량관측소의 경우 국토해양부(437개소), 한국수자원공사(160개소), 기상청(541개소(중관용: 77, 방재용: 464)) 등이 설치되어 있고 수위관측소는 국토해양부(331개소), 수자원공사(112개소) 등이 설치되어 있다. 그러나 대부분의 유역들은 정확하고 미결측된 강우자료를 획득할 수 있는 반면에 유역출구점에서의 장기간의 연속적인 시/일 유량자료를 확보하기가 어려운 실정이다. 또한 이와 같은 자료구축 상황에서 유출모의를 통한 유역의 수문성분을 모의하는 것은 더욱 어렵다고 하겠다.

유출모의 목적이 미계측유역의 유출해석 또는 관측이 어려운 수문성분을 추정하기 위함이라고 할 때, 미계측유역에 대한 합리적인 유출모의 방법을 찾는 것은 우리나라뿐만 아니라 전세계적으로 지속적인 관심을 가지는 분야이다. 이와 관련해서 가장 많은 연구가 이루어지고 있는 분야가 '지역화' 이다. 지역화는 모형 적용 시 필요한 매개변수를 계측유역에서 미계측유역으로 전이, 강우인자와 지형인자를 이용한 기후대 구분, 계측유역의 홍수량자료를 이용한 미계측유역의 홍수해석, 수문학적 동질성을 갖는 유역구분 등에 활용되고 있다.

지역화와 관련된 국내 적용사례로 황준식 등(2007)은 금강유역에 대해 물수지모형인 abcd모형(Thomas, 1981)의 매개변수를 회귀식을 이용하여 지역화하였으며, 배덕효 등(2007)과 정일원 등(2007)은 PRMS 모형의 8개 매개변수를 6개 댐유역에서 산정하고 우리나라 전역의 139개 단위유역에 대해 전이하여 유출모의를 수행하였다. 또한, 이상호와 강신욱(2007)은 15개 계측유역에 대해 탱크모형의 9개 매개변수를 추정하고 회귀식을 이용하여 미계측유역에 전이한 사례가 있으며, 남우성 등(2008)은 요인분석과 군집분석을 이용하여 수문학적으로 동질한 지역을 구분하여 지역빈도해석

에 이용한 사례가 있다. 국외 적용사례의 경우 Zhang and Hall(2004)은 중국 Gan-Ming River 유역에 지역빈도해석을 위해 86개 수위관측소에 대해 4개의 군집방법(Residuals method, Ward's cluster method, Fuzzy c-means method and Kohonen neural network)을 적용하였다. Rao and Srinivas(2006)는 합성군집방법(hybrid-clustering method)을 미국 Indiana 지역에 적용하고 4가지 평가지표를 이용하여 군집결과의 타당성을 평가하였다. 이효상과 문영일(2007)은 회귀분석을 통해 계측유역의 매개변수를 미계측유역으로 전이시키는 방법을 이용하여 영국 126개 유역에 대해 강우-유출 모형의 적용 및 지역화 검정을 수행하였으며, Quarda *et al.*(2008)은 계층적 군집방법(hierarchical clustering method)과 Canonical 상관분석법을 멕시코 3개 하천에 적용하여 지역홍수빈도해석을 수행하였다. 또한 Cloutier *et al.*(2008)은 계층적 군집방법과 주성분분석(principal component analysis)를 이용하여 지형자료와 지하수의 관계규명을 통해 지하수의 공간적인 분포를 분석하였다.

모형의 매개변수 지역화 방법은 적용하고자 하는 모형의 기본구조(개념적, 물리적)와 형태(집중형, 분포형), 모형의 계산시간과 분석기간, 계측자료의 활용 정도, 미계측유역의 자료추출 상황 등을 종합적으로 고려하여 선택하여야 한다.

본 연구에서는 준분포형 유출모형인 SWAT(Soil and Water Assessment Tool) 모형을 우리나라 109개 중권역 유역에 적용할 경우 미계측유역의 유출모의를 위한 매개변수 지역화 방법을 제안하고자 한다. 이를 위해 모형 변수 중 추정이 어려운 매개변수와 관계된 유역 및 수문특성인자를 추출하고 주성분분석과 군집분석을 결합한 다변량분석을 이용하여 미계측유역을 계측유역으로 군집화하였다. 또한 계측유역의 매개변수를 산정하기 위해 6개 댐유역에 대해 SWAT 모형을 구축하여 매개변수 검·보정을 수행하였으며 지역화 방법을 통해 미계측유역의 매개변수를 결정하였다. 마지막으로 독립된 3개 댐유역에 전이된 매개변수 적용 및 유출분석을 통해 본 연구에서 제안한 지역화 방법의 적용성을 평가하였다.

2. 다변량 통계분석을 이용한 매개변수 지역화

본 연구에서는 매개변수 지역화를 위해 다변량 통계분석을 이용하였으며 적용방법은 Fig. 1과 같다. 우선 유출모형의 매개변수 중 미계측유역으로 전이하고자 하는 매개변수를 결정하고 매개변수들과 상관성이 높은 유역/수문인자를 추출한다. 이들 인자는 군집분석에 적용될 설명변수를 산정하기 위해 사용된다. 서로 독립된 설명변수 적용과 군집분석의 효율성을 높이기 위해 주성분분석을 이용하여 산정된 성분점수(component score)를 설명변수로 채택하고 이들 설명변수로부터 군집분석을 이용하여 계측유역과 미계측유역간의 유사성 분석을 통해 미계측유역을 계측유역으로 분류하게 된다. 본 연구에서 적용한 주성분분석과 군집분석의 이론을 기술하면 다음과 같다.

Fig. 1. Flow Chart of Parameter Regionalization

2.1 주성분분석

주성분분석은 고차원의 자료를 저차원의 자료로 축소하거나 복잡한 고차원의 자료를 해석하기 위한 다변량 자료분석 기법이다. 일반적으로 다양한 특성을 가지는 자료 또는 현상을 해석하기 위해 사용되거나 재생산 또는 축소된 자료는 다른 통계기법에 적용될 수 있다. 주성분분석은 평균과 분산의 통계적 성질을 이용한 기법으로 자료를 설명하기 위해 투영시켜 분산이 최대가 되는 기본벡터를 찾게 되고, 이때 분산의 크기는 기본벡터가 가질 수 있는 자료의 정보 양을 의미한다. 이들 기본벡터 중 서로 직교하는 성질을 만족하는 기본벡터를 주성분이라고 하고 이 벡터에 해당하는 성분점수를 새로운 자료로 이용하게 된다. 주성분분석의 해석과정을 간단하게 설명하면 다음과 같다.

주성분분석은 선형모형으로 나타낼 수 있으며 일반적인 형태는 Eq. (1)과 같다. Eq. (1)에서 주성분점수로 이루어진 행렬 PC는 각 관측단위와 주성분들의 관계 정도를 나타낸다. X는 $(p \times 1)$ 을 가지는 기존자료의

행렬, v 는 고유값(eigenvalue)에 대한 고유벡터(eigenvector)를 의미한다. 고유값은 Eq. (2)와 같이 각 PC의 분산으로부터 산정된다. 또한 주성분은 상호 독립이기 때문에 i 와 k 가 다를 경우 Eq. (3)과 같이 공분산은 0의 값을 가지고 공분산행렬은 고유값 $\lambda_1 \geq \lambda_2 \cdots \lambda_p \geq 0$ 로 이루어지게 된다.

여기서 i 번째 주성분은 다음의 3가지 제약조건하에서 분산을 최대화하는 선형결합으로 나타낼 수 있다. 1) 분산은 선형결합계수로 사용되는 고유벡터 v 에 따라 증가될 수 있기 때문에 요소들의 자승합이 1이 되도록 제한한다; 2) 주성분분석에서 두 개의 선형결합 관계는 서로 독립이라고 가정하기 때문에 두 주성분의 공분산은 0이 되어야 한다; 3) 자료에서 최대의 고유값을 산출하기 위해 특성방정식(characteristic equation) $|A - \lambda I| = 0$ 이 성립하여야 한다. 여기서 A 는 기존자료의 상관계수행렬 또는 표준화된 변수의 분산-공분산 행렬, I 는 단위행렬을 의미한다.

$$PC = v' \cdot X \quad (1.a)$$

$$PC_1 = v_1' X = v_{11}X_1 + v_{21}X_2 + \cdots + v_{p1}X_p \quad (1.b)$$

$$PC_2 = v_2' X = v_{12}X_1 + v_{22}X_2 + \cdots + v_{p2}X_p$$

$$\vdots$$

$$PC_p = v_p' X = v_{1p}X_1 + v_{2p}X_2 + \cdots + v_{pp}X_p$$

$$\text{Var}(PC_i) = v_i' \Sigma v_i = \lambda_i; \quad i = 1, 2, \dots, p \quad (2)$$

$$\text{Cov}(PC_i, PC_k) = v_i' \Sigma v_k = 0; \quad i \neq k \quad (3)$$

2.2 군집화 방법

객체(object)들을 군집화 하는 방법은 자료의 특성, 군집방법의 특성, 도출하고자 하는 결과 등에 따라 다양하다. 대부분의 방법은 군집내의 객체간 유사성 거리는 되도록 가깝게 하고 군집간 유사성 거리는 가능한 멀게 하는 것이 기본원리이다. 군집화 방법은 계층적 방법과 비계층적 방법으로 구분할 수 있다. 계층적 방법에는 각 군집간의 유사성을 계산하는 방법에 따라 최단연결법, 최장연결법, 중심연결법, 중위수연결법, 군집간 평균연결법, 군집내 평균연결법, Ward 방법 등이 있고 비계층적 방법은 K-평균법이 대표적인 방법이다. 계층적 방법은 지구과학분야(Davis, 1986)와 수문지리분야(Steinhorst and Williams, 1985; Schot and van der Wal, 1992; Ribeiro and Macedo, 1995; Güler *et al.*, 2002; Cloutier *et al.*, 2008)에서 수문학적 동질성 분석을 위해 널리 활용되고 있는 방법이다. 또한 최근에는

각 군집화 방법의 장·단점을 보완하기 위해 합성 군집화 방법(hybrid cluster analysis)을 적용하는 사례(Rao and Srinivas, 2006)가 있다.

일반적으로 이러한 방법들은 유역 또는 관측소에 대해 홍수, 강수, 기후 등의 수문기상학적 동질성을 갖는 유역들로 나누기 위해 사용되는 방법이다. 본 연구에서는 준분포형 유출모형의 매개변수를 미세측유역으로 전이하기 위한 방법으로 앞서 언급한 방법을 직접적으로 적용하기는 어렵다. 따라서 계층적 군집방법 중 최단연결법을 응용하여 적용하였다. 기존 최단연결법은 유사성거리가 가장 가까운 객체들을 묶는 것을 시작으로 순차적으로 유사성거리가 가까운 객체 또는 객체와 집단을 서로 군집시키는 방법이다. 본 연구에서는 계층유역을 군집의 초기 객체로 하여 모든 미세측유역과 유클리드 거리를 계산하고 계층유역과 유사성거리가 가장 가까운 미세측유역을 해당 계층유역에 군집되도록 하였다. 적용 식은 Eq. (4)와 Eq. (5)이다. 식에서 X_i 는 i 번째 계층유역, Y_j 는 j 번째 미세측유역, $C\{X_i\}$ 는 i 번째 계층유역 군집, $d(X_i, Y_j)$ 는 Eq. (5)로부터 산정된 i 번째 계층유역과 j 번째 미세측유역간의 유사성거리, x, y 는 각각 계층유역과 미세측유역의 설명변수, n 은 설명변수의 개수를 의미한다.

$$C\{X_i\} = \min\{d(X_i, Y_j)\} \quad (4)$$

$$d(X_i, Y_j) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (5)$$

3. SWAT 모형 구축

본 연구에서는 유출분석을 위해 SWAT 모형을 적용하였다. SWAT 모형은 미국 농무성 농업연구소(ARS)에서 개발된 모형으로 기상, 토양형태, 토지이용, 토지피복, 지형 및 지질자료 등에 대한 정보를 바탕으로 광범위한 지역에 대해 유출, 유사, 수질 등의 복합적인 수자원 관리를 할 수 있는 준 분포형 모형(Neitsch *et al.*, 2001)으로 최근 우리나라에서도 다양한 분야에서 활발하게 적용되고 있다. 모형의 최소 계산단위는 수문학적 반응단위(Hydrologic Response Unit; 이하 HRU)이며 이는 같은 유역에서 동일한 토지피복 및 토양특성을 갖는 지역으로 정의할 수 있다. Fig. 2는 HRU 단위의 토양층에 대한 유출수문성분 거동을 나타낸 것이다. SWAT 모형은 토양층을 10개로 구성할 수 있으며 강수(PCP)에서 지표면유출(Q_{sur})을 제외한 물은 토양으로 침투되고, 각 토양층의 포장용수량(field capacity,

FC)와 영구위조점(wilting point, WP)을 기준으로 자유수(free water)와 장력수(tension water)로 구분되어 증발호름(Q_{lat}), 침투(Q_{sep}), 증발산량(ET) 형태로 이동하게 된다. 또한 최하부 토양층에서의 침투량은 대수층(SA, DA)의 함양량으로 전환되고 이 양은 지하수유출(Q_{gw})을 형성하게 된다. 모든 수문성분 발생량은 각 HRU별로 모의되고 이는 해당 소유역의 주하도로 유입된다. 또한 하도에서 합쳐진 양은 각각 계산단위시간(Δt)동안 주하도의 하도추적을 통해 해당 소유역 밖으로 유출된다(이병주 등, 2008).

Fig. 2. Schematic Representation of the Soil Layer Structure of SWAT

SWAT 모형 적용 시, 기존 모형은 계산시간이 일단 위일 경우 하도추적 과정에서 불안정한 유출결과가 모의되는 단점이 있다. 김남원 등(2007)은 이러한 문제점을 개선하기 위해 비선형 저류함수모형을 이용한 하도추적방법을 SWAT 모형에 적용한 사례가 있으며 본 연구에서도 이 방법을 적용하였다.

본 연구에서는 제주도를 제외한 우리나라 중권역을 대상으로 SWAT 모형을 구축하였다. 중권역은 대권역의 분류를 기준으로 큰 지류하천이 자연적으로 합류하는 지점을 기준으로 분할되었으며, 한강 26개, 낙동강 33개, 금강 21개, 섬진강 15개, 영산강 14개의 총 109개 유역으로 구성되어 있다. 대상유역의 유역특성인자는 국토해양부에서 제공하는 DEM(30m×30m), 토지피복도(30m×30m)를 이용하여 추출하였으며 농업과학기술원으로부터 정밀토양도(1:25000)와 토양 속성자료를 제공받아 토양특성인자를 산정하였다.

4. 중권역 유역 지역화

4.1 모형변수 선정

설명변수를 추출하기 위해서는 지역화를 통해 얻고자 하는 것이 무엇인지를 명확히 규명하는 것이 중요하다

다. 본 연구에서 추정하고자 하는 모형변수는 유출결과에 민감하며 직·간접적인 계측 또는 추정이 어려운 변수들이다. 본 연구의 유출해석에 이용된 SWAT 모형은 물리적 기반의 준분포형 강우-유출모형으로 상당수의 모형변수는 DEM, 토지피복도, 토양도 등의 지형공간자료와 기존에 연구된 식생과 토양특성자료로부터 산정된다. Table 1은 지역화를 통해 추정하고자 하는 모형변수를 나타낸 것이다. ESCO는 토양증발보상계수로서 실제증발산량에 직접적인 영향을 미치고 이로 인해 토양층 내 장력수에 해당하는 토양수분의 감소량이 결정된다. SLSOIL은 경사장으로 토양층내 자유수의 흐름에 영향을 줌으로 수문성분 중 측방흐름 즉, 중간유출량과 직접적인 관계가 있다. GW_DELAY는 토양수분이 마지막 토양층에서 침투되어 대수층의 지하수 함양량으로 이동하는데 걸리는 시간을 의미하며 지하수유출의 지체에 영향을 준다. 마지막으로 CH_N2는 주하도의 조도계수로서 하도의 흐름을 모의하는데 영향을 준다.

4.2 유역/수문인자 추출

Table 1의 4개 모형변수는 특정 유역변수 하나에 영향을 받기 보다는 유역규모, 토지피복, 토양특성 등에 의해 복합적인 영향을 받는다. 따라서 이러한 특성을 반영할 수 있는 유역특성변수로 소유역면적, 평균표고, 평균경사, 산림면적비, 최대토양수분량, 포장용수량, 영구위조점을 선정하였다. Fig. 3은 우리나라 109개 중권역에 대한 변수를 산정한 결과이다. 한강과 낙동강과 같은 상대적으로 큰 대권역 이내의 중권역 유역면적은

크고, 반대로 금강, 섬진강, 영산강 권역 이내의 중권역 유역면적은 상대적으로 작은 것으로 나타났다. 유역평균표고는 태백산맥이 형성되어 있는 평화의 댐, 소양강 댐, 충주댐 유역에서 가장 높게 나타났다. 유역평균경사는 태백산맥이 위치해 있는 지역이 상대적으로 경사가 급한 것을 확인할 수 있고 산림면적비 또한 이 지역에서 높은 것으로 나타났다. 최대토양수분, 포장용수량, 영구위조점은 토양층이 얇게 형성되어 있는 상류유역에 비해 바다와 인접한 유역하류부분에서 많이 분포하는 것을 확인할 수 있다.

4.3 주성분분석을 이용한 설명변수 산정

설명변수는 특수한 경우를 제외하고 서로 독립적이어야 한다. 서로 유사한 상관성을 가지는 다수의 설명변수를 채택할 경우 군집분석의 결과는 왜곡될 수 있다. 앞서 산정된 7개 인자의 상관관계를 분석한 결과 Table 2와 같이 나타났다. 유역면적(V1)은 다른 설명변수들과 상관관계가 낮으며, 평균표고(V2), 평균경사(V3), 산림면적비(V4)와 최대토양수분량(V5), 포장용수량(V6), 영구위조점(V7)간의 상관계수가 각각 높게 나타남을 확인할 수 있다. 따라서 본 연구에서는 7개의 변수를 군집분석의 설명변수로 모두 사용하기 보다는 주성분분석을 수행하여 성분점수를 군집분석의 설명변수로 이용하였다. 주성분 분석 결과 제1, 2 성분의 누가분산비는 82.11%로 선정되었으며 누가분산비가 통상 80%를 넘는 성분까지 취한다는 경험적 기준에 의해 제

Table 1. Model Parameters

Model Parameter	Definition (unit)	Value		
		Initial	Minimum	Maximum
ESCO	Soil evaporation compensation coefficient	0.95	0.0	1.0
SLSOIL	Slope length for lateral subsurface flow (m)	50	0.0	100.0
GW_DELAY	Delay time for aquifer recharge (days)	31	0.0	500.0
CH_N2	Manning's "n" value for the main channel	0.014	0.0	0.3

Table 2. Correlation Coefficient Between Each 7 Variables

Variable	V1	V2	V3	V4	V5	V6	V7
V1	1.00						
V2	0.45	1.00					
V3	0.24	0.84	1.00				
V4	0.22	0.74	0.89	1.00			
V5	-0.05	-0.57	-0.75	-0.61	1.00		
V6	0.04	-0.32	-0.53	-0.42	0.89	1.00	
V7	-0.13	-0.64	-0.69	-0.59	0.84	0.71	1.00

Fig. 3. Estimation of the Characteristics on the Sub-basins for Applying Cluster Analysis

Fig. 4. Plot of Loadings for the First Two Components

1주성분과 제2주성분을 채택하였다.

Fig. 4는 두 성분 축에 대한 7개 변수의 상관관계를 도시한 것이다. 그림으로부터 제1주성분 축에 대해 평

균표고, 평균경사, 산림면적비는 + 상관성을 보이고 최대토양수분량, 포장용수량, 영구위조점은 - 상관성을 보이는 것을 확인할 수 있다. 일반적으로 상류 산지 지형을 많이 포함하는 유역은 토양층이 얇기 때문에 최대토양수분량, 포장용수량, 영구위조점이 상대적으로 적게 나타나고 유역평균표고는 높고, 평균경사가 크며 산림면적비가 큰을 고려할 때 제1성분은 유역의 상하류 위치와 높은 상관성이 있다고 하겠다. 제2성분 축은 유역면적과 평균표고, 평균경사, 산림면적비가 + 상관성을 보이는 것으로 나타나므로 유역크기와 밀접한 관계가 있음을 예상할 수 있다.

4.4 군집분석 수행

군집분석에 적용될 두 성분 축의 성분점수를 도시한 결과는 Fig. 5와 같다. 앞서 분석된 성분 축의 의미를 고려할 때 전체적으로 중권역 유역들의 유역크기와 상하류 위치가 다양하게 분포하는 것을 확인할 수 있다. 다만, 대권역별로는 섬진강유역과 영산강유역에 속한

Fig. 5. Factor Score of Mid-sized Sub-basins

중권역들이 유역면적이 작고 상대적으로 하류에 위치한 유역이 많은 것을 확인할 수 있다.

계측유역이 전체 미계측 유역의 특성을 반영할 수 있도록 다양하게 분포되어 있다면 군집결과에 대해 보다 높은 신뢰성을 확보할 수 있으나 본 연구에서 분석 결과 계측 가능한 지역이 6개의 댐 지점에 불과하므로 대체로 유역면적이 크고 상류의 산지지역에 편중되는 결과를 보인다. 향후 이러한 부분은 많은 계측지역을 확보함으로써 해결해 나갈 수 있을 것으로 판단된다.

본 연구에서는 Fig. 5의 점들에 대해서 괴산, 안동, 임하, 합천, 용담, 섬진강댐의 6개 계측유역(Dam site)과 103개 미계측 유역간의 유클리드 거리(Eq. (5))를 계산하여 미계측 유역을 거리가 가장 가까운 계측유역에 포함시켰다.

Fig. 6은 103개 미계측유역을 6개의 계측유역으로 분류한 결과이다. 섬진강댐이 44개로 가장 많은 미계측유역을 포함하며 괴산댐 23개, 합천댐 21개, 안동댐 6개, 임하댐 5개, 용담댐 4개 순으로 나타났다. 한강, 금강,

영산강 하류지역은 괴산댐에 포함되고, 충주댐과 남강댐 상류유역과 화천댐과 소양강댐 직상류유역은 안동댐에 포함되는 것으로 나타났다. 또한 합천댐에는 주로 낙동강과 섬진강의 다수 유역이 포함되고, 임하댐과 용담댐에는 동해안 유역을 포함한 일부 유역이 포함되는 것으로 분석되었다.

본 연구에서는 군집분석결과에 대한 정확도를 평가하기 위해 분산분석(analysis of variance; ANOVA)을 수행하였다. 분산분석은 집단내 변동(within groups variability)과 집단간 변동(between groups variability)

Fig. 6. The Result of Cluster Analysis for Parameter Regionalization

Table 3. ANOVA Table of the 6 Groups on the Scores of the Component 1st and 2nd

component Score	Cause of Variability	Sum of Squares	dof	Mean Square	F
1st	Between Groups	78.07	5	15.61	53.72
	Within Groups	29.93	103	0.29	—
	Total	108.00	108	—	—
2nd	Between Groups	67.87	5	13.57	34.83
	Within Groups	40.13	103	0.39	—
	Total	108.00	108	—	—

을 비교하여 집단의 동일성을 검정하는 방법이다. 6개 군집 결과에 대해서 분산분석을 수행하였으며 결과는 Table 3과 같다. 가설검정을 위한 검정통계량은 ‘집단 내’에 대한 ‘집단간’의 평균제곱비로 산정되며 Table 3에서 F값에 해당한다. 1% 유의수준에서 임계값 F(5, 120, 0.01)는 3.17로 제1, 2 주성분으로부터 산정된 F값 53.72와 34.83은 임계값보다 훨씬 큰 값을 갖는다. 따라서 “집단간 평균이 같다”라는 귀무가설은 기각되고 “집단간 평균의 차이가 존재한다”는 것을 통계적으로 증명될 수 있다.

5. 지역화 정확도 평가

5.1 계측유역 검증

SWAT 모형의 매개변수를 미계측유역으로 전이하기 위해 6개 계측유역에 대한 유출분석을 수행하였다. 보정기간동안 모의유량이 관측유량자료에 근접하도록 매개변수를 보정하였으며, 검정기간에 대해서 추정된 매개변수의 적합도를 평가하였다. 계측유역에 대해 추정된 매개변수 결과는 Table 4와 같다.

Table 5는 6개 계측유역에 대한 모의결과의 통계치를 제시한 것이다. 섬진강댐은 관측유량에 비해 과대추정되어 정확도가 낮게 산정되었으나 나머지 댐은 보정기간과 검정기간에 대해 유출용적오차(VE) 10% 이내,

평균제곱근오차(RMSE) 5mm/d 이내, 상관계수(CC) 0.8 이상, 모형효율성계수(ME) 0.6 이상으로 높은 적합도를 보이는 것으로 나타났다. 또한 유출모의결과에 대한 도식적 평가를 위해 검·보정 기간에 대해 관측-모의 유출수문곡선을 비교하였다. Fig. 7은 전체 모의기간 중 최근 5년의 기간에 대해 일단위 관측-모의유출량을 Hogue *et al.*(2000)가 제안한 변환유량(transformed flow)으로 변환하여 도시한 것이다. 섬진강댐의 경우 저수부에서 모의유량이 다소 크게 모의되기는 하나 전체적으로 6개 댐의 고수부와 저수부에서 모의치가 관측치와 유사하게 거동하는 것을 확인할 수 있다. 이상의 통계적, 도식적 평가로부터 추정된 매개변수는 미계측 유역에 적용되기 위한 충분한 정확도를 가진다고 판단된다.

Table 4. Estimated Parameter on Each Basin

Dam Site	ESCO	SLSOIL	GW_DELAY	CH_N2
Goesan	0.85	7.0	90	0.056
Andong	0.90	10.0	90	0.042
Imha	0.85	10.0	85	0.045
Hapcheon	0.83	9.7	80	0.039
Yongdam	0.95	10.5	50	0.040
Seomjin	0.81	10.2	60	0.038

Table 5. Statistical Analysis on Simulated Discharges for Each Dam Site

Dam Site	Calibration/Verification (Period)	Statistics			
		VE (%)	RMSE (mm/d)	CC (-)	ME (-)
Goesan	Cal. (1996 - 2005)	0.4	3.85	0.80	0.64
	Ver. (1982 - 1995)	-4.09	3.35	0.83	0.68
Andong	Cal. (1996 - 2005)	1.06	2.77	0.92	0.81
	Ver. (1977 - 1995)	-4.09	2.00	0.89	0.79
Imha	Cal. (1999 - 2005)	-1.38	3.50	0.87	0.76
	Ver. (1993 - 1998)	6.65	2.75	0.84	0.69
Hapcheon	Cal. (1999 - 2005)	-2.19	2.86	0.92	0.83
	Ver. (1989 - 1998)	5.53	2.15	0.91	0.80
Yongdam	Cal. (2003 - 2005)	0.05	4.48	0.90	0.77
	Ver. (2001 - 2002)	-8.82	3.11	0.89	0.80
Seomjin	Cal. (1996 - 2005)	10.03	5.36	0.73	0.52
	Ver. (1975 - 1995)	21.38	4.95	0.62	0.30

Correlation coefficient (CORR) = $SS_{os} / \sqrt{SS_o \times SS_s}$, $SS_{os} = \sum (O_i - \bar{O})(S_i - \bar{S})$, $SS_o = \sum (O_i - \bar{O})^2$, $SS_s = \sum (S_i - \bar{S})^2$

Root mean square error (RMSE) = $\sqrt{\sum (O_i - S_i)^2 / n}$, where, n is a number of data.

Nash-Sutcliffe efficiency (ME) = $[\sum (O_i - \bar{O})^2 - \sum (O_i - S_i)^2] / \sum (O_i - \bar{O})^2$

Percent error in volume (VE) = $100 \times (\sum S_i - \sum O_i) / \sum O_i$

where, O is observed flow and S is simulated flow.

Fig. 7. Observed and Simulated Discharges for 6 Gauged Dam Sites

5.2 미계측유역 검정

계측유역의 매개변수를 미계측유역으로 전이하였을 경우 유출모의결과의 정확도를 분석하기 위해 소양강댐, 충주댐, 대청댐의 3개 댐 지점을 미계측유역으로 가정하여 매개변수 전이 및 유출모의를 수행하였다. 댐 상류 중권역 유역의 매개변수는 Fig. 6에서 충주댐(1001-1003), 소양강댐(1011, 1012), 대청댐(3001-3008)의 계측유역에 해당하는 매개변수를 전이하여 모의하였

다. 대청댐의 모의기간은 댐 상류에 위치한 용담댐의 효과를 배제하기 위해 용담댐 준공년도인 2001년 이전까지의 기간을 모의기간으로 적용하였다. Table 6은 각 댐의 모의기간과 유출모의결과에 대한 통계치를 제시한 것이다. 유출용적오차에서 다소 차이는 있으나 나머지 통계치에 대해서는 상당히 높은 정확도를 보이는 것으로 나타났다. Fig. 8은 전체 모의기간 중 최근 5년의 기간에 대해 일유출량 모의결과와 관측유량을 변환유량으로 표시한 것이다. 3개 댐 모두 고수 및 저수유출량 모

Fig. 8. Graphical Analysis of Simulated Discharges for the Verification of Regionalized Param

Table 6. Statistical Analysis of Simulated Discharges Using Regionalized Parameters

Dam Site (Period)	VE (%)	RMSE (mm/d)	CC (-)	ME (-)
Soyang (1974 - 2005)	-15.50	2.76	0.94	0.85
Chungju (1986 - 2005)	4.82	2.98	0.90	0.81
Daechong (1981 - 2000)	13.32	2.33	0.90	0.80

의에서 관측치에 근접하게 모의하는 것을 볼 수 있다.

이상의 결과로부터 계측유역의 매개변수를 미계측유역으로 전이하였을 경우 유출량의 정확도를 간접적으로 평가하였으며 통계적 평가와 도식적 평가에서 모두 적합도가 높게 나타남을 확인할 수 있었다.

6. 결 론

본 연구는 준분포형 유출모형인 SWAT 모형을 우리나라 109개 중권역에 적용할 경우 미계측유역에 대한 매개변수 결정방법을 제안하고 적용 타당성을 검증하고자 하였다. 본 연구에서 도출된 결과를 요약하면 다음과 같다.

- 1) 계측유역의 매개변수를 미계측유역으로 전이하기 위한 방법으로 주성분분석과 군집분석을 결합한 다변량 통계분석을 수행하였다. 설명변수는 주성

분분석을 통해 산정된 성분점수를 이용하였으며 계측유역과 미계측유역간의 유클리드 거리를 산정하여 군집분석을 수행하였으며 군집방법은 계층적 군집방법 중 최단거리법을 수정하여 적용하였다.

- 2) SWAT 모형변수 중 추정이 어려운 4개 매개변수와 밀접한 관계가 있다고 판단되는 유역면적, 평균표고, 평균경사, 산림면적비, 최대토양수분량, 포장용수량, 영구위조점의 7개 유역/수문변수를 109개 중권역에 대해 산정하였다. 대체로 내륙에 위치한 상류 산지유역일수록 유역면적이 크고, 평균표고가 높으며 경사가 급한 것으로 나타났고, 최대토양수분, 포장용수량, 영구위조점은 서해 하류 지역에서 큰 값을 가지는 것으로 분석되었다.
- 3) 주성분 분석 결과 제1성분과 제2성분은 전체 자료의 82.11%를 함축하고 있으며 제1성분은 유역 위치, 제2성분은 유역 규모와 관계가 있는 것으로 분석되었다. 이들 성분점수를 이용하여 103개 미계측유역을 6개 계측유역으로 군집화 한 결과 섬진강댐이 44개로 가장 많은 미계측유역을 포함하고 순위별로 괴산댐 23개, 합천댐 21개, 안동댐 6개, 임하댐 5개, 용담댐 4개 순으로 나타났다. 이들 군집결과에 대해 분산분석기법을 이용하여 정확도를 평가한 결과 군집된 집단의 평균이 서로 상이함을 통계적으로 증명하였다.

4) 6개 계측유역에 대해 SWAT 모형을 적용하여 매개변수를 추정 및 검증하였으며 전이된 매개변수의 적용성을 평가하기 위해 미계측유역으로 가정한 소양, 충주, 대청댐에 대해 추정된 매개변수를 전이하여 유출모의를 수행한 결과 유출용적오차와 평균제곱근오차는 15%와 3mm/d 이내, 상관계수와 모형효율성계수는 각각 0.9와 0.8 이상의 높은 정확도를 갖는 것으로 분석되었다.

본 연구의 결과는 미계측 유역에 대한 유출모형의 적용뿐만 아니라 장기간의 시공간적인 수문성분의 변화를 분석할 수 있는 기초자료로 활용될 수 있다는 측면에서 충분한 가치가 있다고 하겠다. 다만, 지형적으로 유사한 특성을 가지는 6개 댐 지점을 계측유역으로 적용하였기 때문에 다양한 미계측유역의 특성을 반영하기에는 무리가 있을 것으로 판단된다. 향후 다양한 유역/수문특성을 가지는 계측유역을 추가하여 분석한다면 보다 신뢰성 있는 매개변수 지역화 및 유출결과를 도출할 수 있을 것으로 전망된다.

감사의 글

이 연구는 소방방재청 자연재해저감기술개발사업(지자체별 실시간 내배수재해 예측기술 개발) 연구비 지원으로 수행되었으며 이에 감사드립니다.

참 고 문 헌

김남원, 이정우, 이병주, 이정은 (2007). “비선형 저류방정식을 이용한 일 단위 하도추적법.” **대한토목학회 논문집**, 대한토목학회, 제27권, 제5B호, pp. 533-542.

남우성, 김태순, 신주영, 허준행 (2008). “다변량 분석 기법을 활용한 강우 지역빈도해석.” **한국수자원학회 논문집**, 한국수자원학회, 제41권, 제5호, pp. 517-525.

배덕효, 정일원, 권원태 (2007). “수자원에 대한 기후변화 영향평가를 위한 고해상도 시나리오 생성(I): 유역별 기후시나리오 구축.” **한국수자원학회논문집**, 한국수자원학회, 제40권, 제3호, pp. 191-204.

이병주, 배덕효, 김남원 (2008). “SWAT 모형을 이용한 토양수분의 시공간적 변동성 평가.” **한국수자원학회 학술발표회 논문집**, 한국수자원학회, pp. 1270-1274.

이상호, 강신욱 (2007). “유출특성인자를 이용한 수정탱크모형 매개변수의 지역화.” **대한토목학회논문집**, 대한토목학회, 제27권, 제4B호, pp. 379-385.

이효상, 문영일 (2007). “미 계측 지역의 강우 유출 모델의 적용.” **대한토목학회논문집**, 대한토목학회, 제27권, 제5B호, pp. 489-498.

정일원, 배덕효, 임은순 (2007). “수자원에 대한 기후변화 영향평가를 위한 고해상도 시나리오 생성(II): 유역별 유출시나리오 구축.” **한국수자원학회논문집**, 한국수자원학회, 제40권, 제3호, pp. 205-214.

황준식, 정대일, 이재경, 김영오 (2007). “기후변화 영향 평가를 위한 월 물수지모형의 적용성 검토.” **한국수자원학회논문집**, 한국수자원학회, 제40권, 제2호, pp. 147-158.

Cloutier V., Lefebvre, R., Therrien, R., Savard, M.M. (2008). "Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system." *Journal of Hydrology*, Vol. 353, pp. 294-313.

Davis, J.C. (1986). *Statistics and data analysis in Geology*. John Wiley & Sons Inc., New York.

Güler, C., Thyne, G.D., McCray, J.E., Turner, A.K. (2002). "Evaluation of graphical and multivariate statistical methods for classification of water chemistry data." *Hydrogeology Journal*. Vol. 10, pp. 455-474.

Hogue, T.S., Sorooshian, S., Gupta, V.K., Holz, A., Braatz, D. (2000). "A multistep automatic calibration scheme for river forecasting models." *Journal of Hydrometeorology*. Vol. 10, pp. 455-474.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., (2001). *Soil and Water Assessment Tool - Theoretical Documentation(version 2000)*.

Quarda, T.B.M.J., Bâ, K.M., Diaz-Delgado, C., Cârsteanu, A., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E., Bobée, B. (2008). "Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study." *Journal of Hydrology*, Vol. 348, pp. 40-58.

Rao, A.R., Srinivas, V.V (2006). "Regionalization of watersheds by hybrid-cluster analysis." *Journal of Hydrology*, Vol. 318, pp. 37-56.

Ribeiro, L., Macedo, M.E. (1995). "Application of multivariate statistics, trend and cluster analysis to groundwater quality in the Tejo and Sado aquifer." in: groundwater Quality: Remediation and Protection. *Proceeding of the Prague Conference, May 1995. IAHS publ.* No. 225, pp. 39-47.

Schot, P.P., van der Wal, J. (1992). Human impact on regional groundwater composition through

- intervention in natural flow patterns and changes in land use." *Journal of Hydrology*, Vol. 134, pp. 297-313.
- Steinhorst, R.K., Williams, R.E. (1985). "Discrimination of groundwater sources using cluster analysis, MANOVA, canonical analysis and discriminant analysis." *Water Resources Research*, Vol. 21, pp. 1149-1156.
- Thomas, H.A. (1981). Improved Methods for National Water Assessment. Report, WR15249270, U.S. Water Resource Council, Washington, D.C.
- Zhang, J., Hall, M.J. (2004). "Regional flood frequency analysis for the Gan-Ming River basin in China." *Journal of Hydrology*, Vol. 296, pp. 98-117.
- (논문번호:08-102/접수:2008.10.18/심사완료:2009.01.30)