Advances in Water Resources 34 (2011) 865–872

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Regionalization of low flows based on Canonical Correlation Analysis

George Tsakiris, Ioannis Nalbantis *, George Cavadias

Centre for the Assessment of Natural Hazards and Proactive Planning and Laboratory of Reclamation Works and Water Resources Management, School of Rural and Surveying Engineering, National Technical University of Athens, Greece

ARTICLE INFO

Article history:
Received 28 June 2010
Received in revised form 23 January 2011
Accepted 5 April 2011
Available online 12 April 2011

Keywords:
Regional analysis
Low flows
Canonical correlation
Low-flow statistics
Ungauged basins
Jack-knife method

ABSTRACT

Regional analysis of low-flow statistics is a critical step in solving water resources management problems related to the requirements of the Water Framework Directive. Important element in this analysis is the determination of homogeneous sub-regions based on physiographic characteristics of the corresponding basin. The purpose of this paper is to investigate the use of the canonical correlation method for partitioning the set of drainage basins of a region into a number of homogeneous sub-regions and determining the relations between the physiographic and low-flow statistics of the basins of each sub-region. The method is also proposed to be used for classifying an ungauged basin in a sub-region of gauged basins.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Knowledge of low-flow statistics is of great importance in water resources management in many respects. First, these statistics can assist in the design of hydraulic works such as hydropower plants, water diversions for various uses (irrigation, water supply, etc.) water treatment plants and sanitary landfills. Second, they can help in determining operating policies of the above works (e.g., determining minimum downstream flow requirements). Third, they can contribute in determining water quality reference conditions which are required by modern legal frameworks for water resources management such as the Water Framework Directive of the European Union [12].

Yet, the term 'low flow' is perceived in a variety of ways by water resource scientists. According to the definition of the International Glossary of Hydrology [38] low flow is 'flow of water in a stream during prolonged dry weather'. This definition is rather vague and in this work we will keep the term 'low flow' for within-year river flow during the dry season. Thus the distinction between 'low flow' and 'drought' becomes clear and allows us to analyse low flows without reference to drought. The latter is a recurrent natural event, more or less extreme, whose origin is difficult to identify due to complexities in the Global hydrological cycle. Conventionally, less-than-average precipitation over an extended period of time and a specific area results in a so called 'meteorological drought' while also leading to a delayed decrease of river flow or water storage which is termed as 'hydrological drought' [1,19]. Comprehensive reviews have been written for the classification of droughts, the drought indices and the regional dimension of droughts [10,24,23,16,15,19]. This is an active research area within which several European projects have been completed recently (e.g., MEDROPLAN, SEDEMED, PRODIM) producing new approaches for characterising droughts and formulating guidelines for facing short and long term water scarcity problems [31–33].

By adopting streamflow as the key variable for characterizing hydrological drought and by restricting analyses to periods less than a hydrological year it becomes evident that low flow analysis can be an integral part of hydrological drought analysis. In this work analyses are restricted to low flows without reference to drought.

Traditionally, low-flow statistics have been widely used to cope with the natural variability of low flows [37,25]. These are the outcome of low-flow frequency analysis. The most widely used statistic in the United States is the 10-year 7-day average low flow denoted as $Q_{7,10}$ [28]. In general $Q_{d,T}$ denotes the minimum of d-day averages of low flows with a return period of T years or with probability of non-exceedance equal to $(1/T)$. For example $T = 10$ denotes a flow which is exceeded nine years in ten on the average. As an alternative to using isolated flow statistics flow duration curves (FDC) are also used either on a long-term time basis or on the annual basis [3,4]. A kind of intermediate approach consists of using multiple low-flow statistics such as $Q_{7,2}$ and $Q_{7,10}$ employed by Vogel and Kroll [34–37]. These will be used in this work also. Note that the ratio $Q_{7,2}/Q_{7,10}$ is an indicator of the slope of the FDC in its low-flow limb.

* Corresponding author. Tel.: +30 2107722713; fax: +30 2107722632.
E-mail address: nalbant@central.ntua.gr (I. Nalbantis).

0309-1708/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2011.04.007

Regionalization of low-flow statistics has been widely applied in the past with the purpose to estimate low-flow statistics in ungauged basins on the basis of basin physiographic characteristics (geomorphological, climatic, or geological); such studies are reported by Vogel and Kroll [35,37] and Smakhtin [25]. Multivariate regression models have been extensively used to relate low-flow statistics with basin physiographic characteristics. Usually nonlinear multiplicative models are calibrated on data sets from gauged basins to be later applied to any ungauged basin of the area with known values of all explanatory variables. This requirement for explanatory variable information is not always easily fulfilled, e.g., for the case of using the flow recession constant [37].

When using multivariate regression to regionalize low-flow statistics in gauged basins the following questions emerge: (1) Do the scatter diagrams of observed versus estimated low-flow statistics help identify homogeneous sub-regions or, else, basin clusters? (2) If this is true, do the same basin clusters exist in the spaces of both the basin physiographic characteristics and the low-flow statistics? If the answer on question 2 is positive, then how do the regional low flow statistics compare with those based on information from sub-regions only?

When dealing with ungauged basins with known physiographic characteristics a further question arises whether regional analysis can assist in classifying such basins within homogeneous sub-regions before estimating flow variables based on sub-region information only.

To respond to the above questions the method of Canonical Correlation Analysis or CCA [18] will be used which is designed to study the relation between two sets of random variables. For this purpose a well-tested data set from the state of Massachusetts, USA, is used [35,37]. The method has been applied in the past for regional flood estimations [6–9,21,22].

2. Methodology

Assume that the following data are available in matrix form: (1) the basin-related data matrix $\mathbf{X}\,(n \times p)$ containing p characteristics for each one of n gauged basins, and (2) the flow-related data matrix $\mathbf{Q}\,(n \times m)$ (where usually $m \leqslant p$) of the flow-related variables (herein, the low-flow quantiles) for the same set of n gauged basins. Based on these data the canonical correlation coefficients $r_1, r_2, \ldots$, and the two matrices $\mathbf{V}\ (n \times m)$ and $\mathbf{W}\ (n \times m)$ of the corresponding standardized canonical variables $(v_1, v_2, \ldots, v_m)$ and $(w_1, w_2, \ldots, w_m)$ are computed, which are linear combinations of the standardized basin-related and flow-related variables respectively.

By assuming that $m = 2$, the next step is to represent the basins as points in the spaces of the uncorrelated canonical variables (v_1, v_2) and (w_1, w_2) and, then, examine the similarity of the point patterns in these spaces, i.e., the capability of the basin-related canonical variables to predict the flow-related variables. If the two point patterns are sufficiently similar it is examined whether there are sub-groups of points representing basins with different ranges of values of basin physiographic characteristics and flow variables. In this case the basins of different sub-regions belong to different statistical populations and the multivariate regressions mentioned in the introduction of this paper should be carried out separately for each sub-region.

A more detailed description of the method of Canonical Correlation Analysis (CCA) within the context of river flow estimation is given in [6].

3. Existing research results

In an attempt to analyse low flows in the state of Massachusetts Vogel and Kroll studied the relations between the physiographic characteristics and the low flows of 23 basins using a standard multiple linear regression model, after logarithmic transformation of all variables.

They proposed the following multiplicative model for the T-year, 7-day minimum flows of 23 gauging stations in the state of Massachusetts

$$\hat{Q}_{7,T} = b_0 A^{b_1} S^{b_2} K_{\mathrm{b}}^{b_3}, \tag{1}$$

where $Q_{7,T}$ is the 7-day T-year minimum discharge estimated from the streamflow record, $\hat{Q}_{7,T}$ is the corresponding model estimate, A is the watershed area, H is the watershed relief, d denotes the drainage density, S is the average basin slope ($S = 2Hd$), and K_{b} is the base flow recession constant. As stated by Fennessey and Vogel [13] the drainage density, d, is the ratio of the total length of stream channels in the basin, L, divided by the watershed area A. Basin relief, H, is simply a measure of the difference between the basin summit elevation and the channel outlet elevation. So, in this case, A, S and K_{b} are the basin physiographic characteristics and $Q_{7,2}$ and $Q_{7,10}$ are the low-flow statistics. These are the basin-related and the flow-related variables defined in Section 2.

The parameters b_0, b_1, b_2, and b_3 of the model were estimated using the Generalised Least Squares method (GLS) [14,26,27] after a logarithmic transformation of all variables. The GLS method takes into account the cross-correlations and the different record lengths of the flow variables and results generally in smaller model errors in regard to the Ordinary Least Squares (OLS) method. However as pointed out by Vogel and Kroll [35], all sites in Massachusetts had at least ten years of record and the average cross-correlation among concurrent flows was only 0.35 and therefore the GLS procedure led to only marginal gains in the prediction errors as compared to the OLS method which will be used in this study.

Table 1 gives the values of all variables used in the study converted in units of the International System (SI). The coefficients of determination for the multiple regressions of $\ln Q_{7,2}$ and $\ln Q_{7,10}$ on the basin variables were 0.971 and 0.945, respectively.

The scatter diagrams of the observed versus estimated flows (Figs. 1 and 2, adapted from [37]) show two clusters of basins, indicating the existence of two sub-regions in the area studied. Fig. 3 shows the frequency histograms of $\ln Q_{7,2}$ and $\ln Q_{7,10}$ and visually confirms that the flows of the 23 basins are a sample from a mixture of two distributions. As mentioned in Section 1, it would therefore be interesting to examine whether these two clusters correspond to sub-regions with different ranges of values of the basin and flow variables.

The basins of each cluster could be determined using Figs. 1 and 2. However, in order to determine whether these clusters consist of basins with different physiographic and flow characteristics, it is necessary to examine whether the same two clusters exist in the spaces of both the basin and the flow variables. This is achieved through using the CCA method described in Section 2. The data set used by Vogel and Kroll [37], herein called 'raw data set', will be used as the starting point.

4. Application and results

4.1. Model development

Although the raw data set contained data on the watershed relief H, and the drainage density d, Vogel and Kroll preferred including a derived quantity in their analyses which was the basin slope S calculated as $2Hd$ [29,30] (Eq. (1)). In our study we use instead the observed basin variables A, H, d and K_{b} in order to determine their ranges in the basins of each sub-region. The proposed model is therefore

Download English Version:

https://daneshyari.com/en/article/4526104

Download Persian Version:

https://daneshyari.com/article/4526104

Daneshyari.com