Journal of Hydrology 399 (2011) 274–280

Contents lists available at ScienceDirect

Journal of Hydrology

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

Analysis of Sabine river flow data using semiparametric spline modeling

Soutir Bandyopadhyay [a,*], Arnab Maity [b]

[a] Department of Mathematics, Lehigh University, Bethlehem, PA 18015, United States
[b] Department of Statistics, North Carolina State University, Raleigh, NC 27695, United States

ARTICLE INFO

Article history:
Received 10 January 2010
Received in revised form 4 November 2010
Accepted 13 January 2011
Available online 4 February 2011

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Ercan Kahya, Associate Editor

Keywords:
Sabine river
Semiparametric model
Spline

SUMMARY

In this article, a modeling approach for the mean annual flow in different segments of Sabine river, as released in the NHDPlus data in 2007, as a function of five predictor variables is described. Modeling flow is extremely complex and the deterministic flow models are widely used for that purpose. The justification for using these deterministic models comes from the fact that the flow is governed by some explicitly stated physical laws. In contrast, in this article, this complex issue is addressed from a completely statistical point of view. A semiparametric model is proposed to analyze the spatial distribution of the mean annual flow of Sabine river. Semiparametric additive models allow explicit consideration of the linear and nonlinear relations with relevant explanatory variables. We use a conditionally specified Gaussian model for the estimation of the univariate conditional distributions of flow to incorporate auxiliary information and this formulation does not require the target variable to be independent.

1. Introduction

One of the primary challenges for the professionals in water sectors is to meet multiple water demands within the constraint of limited freshwater supply. The necessity to integrate the ecosystem needs is also pronounced in water management. Proper ecosystem management is paramount to protect the ecological processes and biodiversity. It has been noted in some literature that demands for surface water are not expressed freely but rather controlled by water rights specifying the location and type of each allowed usage, the amount to be used and the priority date when the right is established (see for example, http://www.oregonexplorer.info/willamette/). Therefore, a good understanding of available water resources is needed for water professionals to achieve a sustainable water system that enriches both this generation and future, while considering the expected future climate and other relevant geographical and hydrological parameters.

As Mylevaganam and Srinivasan (2008) note, contemporary efforts in planning, designing and implementing resource management efforts are now at the catchment scale. The reason to exploit at the catchment scale is to allow management actions to be carried out unhindered until the magnitude of effect reaches to a point where regulation becomes necessary. It has also been mentioned in Ziemer (1994) that generalized regulations are usually not efficient and usually a higher level of regulation results in more streams being overprotected. The closer that the regulations can be tailored to the variables associated with the risk, the less likely that proposed management actions are curtailed needlessly, or, conversely, the less likely that the regulations are inadequate to protect a desired resource. Added to this, the effect of water resources allocation in the upstream of a river basin plays a crucial role in determining the state of the downstream water availability. The spatial connectivity of stream networks often plays a big role to avoid upstream-downstream conflict. Reliability of a catchment is also indirectly linked to the mean annual flow it conveys.

Further, the availability of hydrological data is also critical for water resources planning. Most drainage basins in this world do not have these data because of poorly developed hydrological networks (Oyebande, 2001; Rodda, 2001). It is also not feasible to establish a flow measuring station on every drainage basin (Chiang et al., 2002) and in addition the sheer sizes of some countries make it impossible to develop adequate hydrological networks and therefore most drainage basins are ungauged (Tucci et al., 1995). Therefore, the need for hydrological data has greatly increased as water resources which are in some cases scarce have to be shared among competing uses.

Therefore, potential to predict water availability, in other words, mean annual flow at a catchment scale considering all the influencing hydrological and geographical parameters is paramount. This also greatly enhances the knowledge on hydrological characteristics of ungauged basins for water resources planning purposes given the prevailing climate and other conditions are of similar nature.

* Corresponding author. Tel.: +1 6107583753.
E-mail address: sob210@lehigh.edu (S. Bandyopadhyay).

0022-1694/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.01.006

1.1. Dataset

In this section we give a brief description of the NHDPlus data. A more detailed description of the NHDPlus can be found in the website of Center for Research in Water Resources (http://www.crwr.utexas.edu/gis/gishydro08/ArcHydro/NHDPlus.htm).

According to NHDPlus Users Guide, NHDPlus (Horizon Systems, 2007) is an integrated suite of application-ready geospatial data products, incorporating many of the best features of the National Hydrography Dataset (NHD), the National Elevation Dataset (NED) and the National Watershed Boundary Dataset (WBD) (Holtschlag, 2009). NHDPlus dataset is distributed for each region as shown in Fig. 1. NHDPlus includes a stream network based on the medium resolution NHD (1:100,000 scale), improved networking, feature naming and "value-added attributes" (VAA). NHDPlus also includes elevation-derived catchments which are produced using a drainage enforcement technique. The VAAs include greatly enhanced capabilities for upstream and downstream analysis and modeling. VAA-based routing techniques are used to produce the NHDPlus cumulative drainage areas and land cover, temperature and precipitation distributions. These cumulative attributes are used to estimate mean annual flow and velocity. The objective of the study is to investigate and propose a lattice based mean annual flow predictor for the NHDPlus dataset as released in 2007.

1.2. Study area and Sabine basin hydrology

Detailed description of origin and flow of Sabine river and its hydrology is provided in Comprehensive Sabine Watershed Management Plan Report (1999), available at the official website of Sabine River Authority of Texas. Below we briefly summarize some of the key points. We refer the interested readers to the original report (located at http://www.sra.dst.tx.us/srwmp/comprehensive_plan/default.asp) for more detailed description of origin, background and hydrology of Sabine river.

Sabine river, a river in the southwestern United States, rises in northeastern Texas, flows southeast and south, broadening near its mouth to form Sabine Lake and continues from Port Arthur through Sabine Pass, a dredged navigable channel, to the Gulf of Mexico after a course of 578 mi (930 km). It drains 10,400 sq mi (26,950 sq km) entirely in Texas and the Louisiana Coastal Plain. The Sabine is a flat-water river that pumps about 6.8 million acre-feet into the Gulf and is the single largest volume river in Texas in terms of its discharge. The water has the tannin acid brown color that is common in East Texas rivers and streams.

The Sabine River Authority of Texas was created by the Legislature in 1949 as an official agency of the State of Texas. The main purpose of this agency was to act as conservation and reclamation district with responsibilities to control, store, preserve and distribute the waters of the Sabine river and its tributary streams for useful purposes. The boundaries were established by the Act of the Legislature and it comprise all of the area lying within the watershed of the Sabine river and its tributary streams within the State of Texas. The watershed area includes all parts of 21 counties. Fig. 2a shows the total number of catchments available in Texas. We consider the data set of catchments only in the Sabine river basin (Fig. 2b) containing 5654 catchments.

The hydrology of Sabine river basin is characterized by diverse climatological, topographical and geological features as well as several climatological factors such as temperature, rainfall and humidity. It is known that topography and geologic factors can affect runoff, evaporation, sedimentation rates, reservoir storage capacity and water quality and define the river system within the basin. As mentioned in Comprehensive Sabine Watershed Management Plan Report (1999), the hydrology of the northern region of the basin is significantly different from the southern region. These distinct regions are commonly referred to as the "Upper basin" in the north and the "Lower basin" in the south, the division between the two areas being the headwaters of Toledo Bend Reservoir. The Upper basin is characterized by cool winters, hot summers and seasonal rainfall patterns. The Lower basin has a coastal climate with mild winters, high annual rainfall and moderate to high humidity. However, in this paper, for the modeling purpose we have not considered these two regions separately. We assume that even if the two regions are distinct from the hydrological point of view but the flow at any catchment can be modeled in the same way for both the regions. The flow at any catchment only depends on a small number of the neighboring catchments and we assume that the hydrological properties of a particular catchment is not significantly different from the properties of its neighboring catchments. The geological factors affect the neighboring catchments similarly in each region, and hence affecting the covariates (precipitation, temperature, etc.) similarly, but not necessarily changing the dependence structure among the neighboring catchments.

In this paper we are going to model the mean annual flow of Sabine river based on the NHDPlus data set released on 2007 in its different catchments based on several relevant variables such as length, stream order, temperature, precipitation and slope. Detailed distributions of these variables based on our data set are shown in Tables 1 and 2.

The article is organized as follows. In Section 2, we discuss our methodology and implement it to model the data. In Section 3 we discuss the implications of the fitted model.

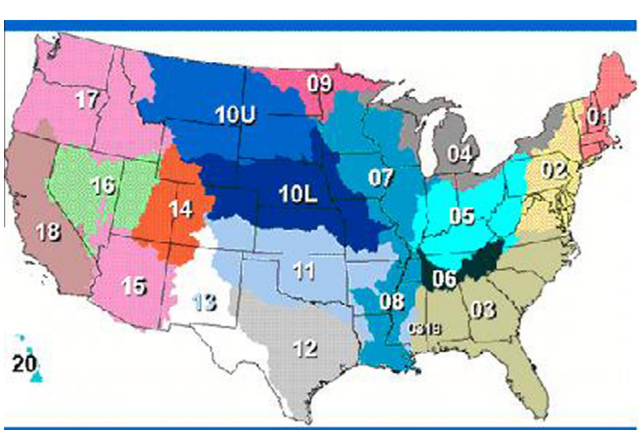

Fig. 1. NHDPlus region.

2. Data analysis

The goal of the analysis performed here and the features of the data at hand give precise indications about the model to be used. First, it is clear that if an event occurs in a region, it is likely to affect the neighboring regions as well, *i.e.*, the events are spatially dependent. The second aim consists in estimating the flow distribution as a function of explanatory variables because flow can be related to a number of factors, for example, precipitation at the specified catchment, temperature, slope of the region, etc. For modeling purpose, the logarithmic transformation of the flow values are considered as a function of relevant explanatory variables. By doing this we implement the constraint that the response variable, flow of the river in a catchment, is always a non-negative quantity.

Complex functional relations characterizing the flow of the river and their spatially dependent structure lead to the adoption of a semiparametric lattice model. In this data we have five covariates,