Journal of Hydrology 533 (2016) 28–39

Contents lists available at ScienceDirect

Journal of Hydrology

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

An integrated statistical and data-driven framework for supporting flood risk analysis under climate change

Y. Lu [a], X.S. Qin [a,*], Y.J. Xie [b]

[a] School of Civil & Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b] Institute of Catastrophe Risk Management (ICRM), Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 19 January 2014
Received in revised form 23 November 2015
Accepted 26 November 2015
Available online 5 December 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Harald Kunstmann, Associate Editor

Keywords:
Climate change
Hydrologic modeling
Downscaling
Flood frequency analysis

SUMMARY

An integrated statistical and data-driven (ISD) framework was proposed for analyzing river flows and flood frequencies in the Duhe River Basin, China, under climate change. The proposed framework involved four major components: (i) a hybrid model based on ASD (Automated regression-based Statistical Downscaling tool) and KNN (K-nearest neighbor) was used for downscaling rainfall and CDEN (Conditional Density Estimate Network) was applied for downscaling minimum temperature and relative humidity from global circulation models (GCMs) to local weather stations; (ii) Bayesian neural network (BNN) was used for simulating monthly river flows based on projected weather information; (iii) KNN was applied for converting monthly flow to daily time series; (iv) Generalized Extreme Value (GEV) distribution was adopted for flood frequency analysis. In this study, the variables from CGCM3 A2 and HadCM3 A2 scenarios were employed as the large-scale predictors. The results indicated that the maximum monthly and annual runoffs would both increase under CGCM3 and HadCM3 A2 emission scenarios at the middle and end of this century. The flood risk in the study area would generally increase with a widening uncertainty range. Compared with traditional approaches, the proposed framework takes the full advantages of a series of statistical and data-driven methods and offers a parsimonious way of projecting flood risks under climatic change conditions.

1. Introduction

A warmer climate may lead to more frequent and intense floods for many areas around the world (IPCC, 2007; Schmocker-Fackela and Naef, 2010; Hirabayashi et al., 2013). This is especially alarming for central south part of China, where the region has already been suffering from continuous flood disasters over many decades. It is thus imperative to gain an in-depth understanding on the future risk of floods at this region, before effective adaptation planning efforts are to be made. The previous research efforts that looked into climate-change impact on hydrological systems were mainly based on integrated climatic and hydrological modeling. The future climate change condition is mainly modeled by general circulation models (GCMs) considering various emission scenarios. Due to resolution problems, GCM model outputs are generally difficult to be used directly for hydrological impact studies, especially for those small to medium size watersheds (Tisseuil et al., 2010). Many researchers have tried to develop either dynamic or statistical downscaling tools to help convert the coarse GCM data into higher-resolution local weather data (Wilby and Wigley, 1997; Wilby et al., 1999; Fowler et al., 2007; Lu and Qin, 2014). Hydrological model is an essential tool to help build the linkage between weather information and river runoff. Previously, a wide range of hydrological models have been developed, and most of them describe the rainfall-runoff transformation processes based on physical theory (Zhao and Liu, 1995; Refsgaard and Storm, 1995; Arnold et al., 1998; HEC, 2000; Huo et al., 2012; Wang et al., 2012; Fu et al., 2013).

From the previous studies, it is found that dynamic downscaling is normally computationally intensive and physically-based hydrological modeling has strict requirement on input data. This poses significant challenges for developing countries where relatively limited resources and data are available (Gao et al., 2010). The statistical and/or data-driven tools, which are computationally cheaper and efficient, have gained their popularity in the past decades in the fields of either downscaling or hydrological modeling (Fowler et al., 2007). For example, in terms of downscaling, the Statistical DownScaling Model (SDSM), Automated regression-based Statistical Downscaling tool (ASD) and Generalized Linear Model (GLM) are the most popularly used tools based on linear regression theory (Wilby et al., 2002; Chandler and Wheater, 2002; Yang

* Corresponding author. Tel.: +65 67905288; fax: +65 67921650.
E-mail address: xsqin@ntu.edu.sg (X.S. Qin).

http://dx.doi.org/10.1016/j.jhydrol.2015.11.041
0022-1694/

et al., 2005; Khan et al., 2006; Dibike et al., 2008; Hessami et al., 2008; Hashmi et al., 2009; Chen et al., 2012). The data-driven approaches, like Artificial Neural Network (ANN) and support vector machine (SVM), are found to be effective in downscaling weather data, especially temperature (Wilby and Wigley, 1997; Schoof and Pryor, 2001; Coulibaly and Dibike, 2005; Khan et al., 2006; Moriondo and Bindi, 2006; Okkan and Serbes, 2012; Liu et al., 2013a, 2013b). Other types of statistical downscaling tools, such as weather generator (Semenov and Barrow, 1997; Semenov et al., 1998) and weather typing scheme (Fowler et al., 2005 and 2007), are also widely used. In hydrological modeling field, applications of ANN can be found in Huang et al. (2004), Gao et al. (2010), Yilmaz et al. (2011), Song et al. (2012), Piotrowski and Napiorkowski (2011 and 2013), and Nourani et al. (2015).

There are also a number of studies that rely primarily on statistical and/or data-driven approaches for hydrological impact assessment. Zarghami et al. (2011) applied a stochastic weather generator (i.e. LARS-WG) and an ANN model to project the monthly stream-flow responses under three emission scenarios of HadCM3 in Iran. The results indicated a notable reduction of monthly runoff at the end of this century. Sachindra et al. (2011) employed the least square support vector machine (LS-SVM) method to downscale GCM output to monthly stream flow directly, without using hydrologic models. The results showed that the proposed method was reasonable in reproducing stream flows in summer and winter, but poor for autumn. Hassan et al. (2012) applied SDSM and ANN to simulate daily stream flows under climate change conditions in Kurau River, Malaysia. The results demonstrated that the hydrological model based on ANN could effectively reflect the variation of monthly stream flow, and capture the mean and low flows well. However, the ANN model notably underestimated the peak values. Liu et al. (2013a) applied a GLM (named GLIMCLIM) and a non-homogeneous hidden Markov model (NHMM) to downscale the daily climate variables in North China Plain (NCP) under different emission scenarios, and assessed the annual runoff responses using Gardner's method (2009). The results showed that the two downscaling methods could generate consistent results, and the annual runoff would decrease in the period of 2081–2099.

Based on the above-mentioned studies, a number of research gaps are identified. Firstly, although there are some studies on using statistical and/or data-driven approaches for hydrological impact study, many of them focused on prediction of monthly or yearly flows. Daily flow prediction is rarely tackled. This is mainly because the forecasting performance of the data-driven approaches (like ANN) for daily extreme data in a long-term time frame (e.g. more than thirty years) was found relatively poor (Sulaiman et al., 2011; Hassan et al., 2012). Secondly, flood risk assessment is an important task for watersheds in central-south part of China as this region has suffered from serious flooding problems. Some previous studies of hydrological predictions under climate change in the Yangtze River Basin, China, were reported (Long et al., 2008; Xu et al., 2011; Liu et al., 2012); but few of them focused on tackling the variation of peak flows or flood frequencies under climate-change conditions. Moreover, many studies generally relied on single downscaling technique such as LARS-WG, ANN or SDSM to generate multiple weather data. In fact, it has been recognized that different downscaling methods could have different levels of performances in allusion to different weather variables. For example, ANN model is found superior in temperature downscaling for certain regions, but poorer for precipitation (Schoof and Pryor, 2001; Coulibaly and Dibike, 2005); ASD is considered suitable for handling rainfall in large river basins like the Yangtze River Basin, China (Guo et al., 2012). It is thus desired to adopt different statistical models for downscaling different types of weather variables in order to take the full advantage of individual methods considering the cross-correlations among multiple variables.

Therefore, this study aims to develop an integrated statistical and data-driven (ISD) framework to investigate the potential impact of climate change on flood frequencies. The Duhe watershed, in the Yangtze River Basin, China will be used as the study case for demonstration. The data-driven method is used to simulate the monthly runoff and statistical downscaling models are used for projection of rainfall, temperature and relative humidity from GCM model predictors. Daily runoff is generated from monthly river runoff based on a disaggregation technique and the extreme flows are analyzed through flood frequency analysis. The study area is characterized by a transitional climate from humid to sub-humid and representative of watersheds in central-south part of China, where flooding risk has been a major concern. The proposed framework mitigates the limitations of the previous studies mentioned above, and offers a computationally cheap alternative of studying flood risks under climate change.

2. Methodology

2.1. System framework

The ISD framework is proposed for analyzing river flows and flood frequencies under climate change. Fig. 1 shows the overall system diagram. There are four major components in the integrated framework: downscaling, hydrologic simulation, disaggregation and flood frequency analysis. Before applying the ISD framework, the hydrological model (i.e. Bayesian Neural Network, BNN) should be trained by the observed meteorological/runoff data to identify suitable meteorological variables and input scheme. The interactions of different components are described as follows:

(1) The statistical downscaling models are applied to produce a number of meteorological variables (e.g. rainfall, temperature and relative humidity) from GCMs at daily scale. The hybrid model based on ASD and K-nearest neighbor (KNN), named ASD-KNN, is applied for multisite rainfall downscaling. In detail, ASD is used to downscale the mean areal rainfall, and the monthly rainfall is calculated by the summation of daily data. Then, KNN is used to spatially disaggregate monthly rainfall to multiple sites, where the spatial correlations among different sites shall be kept. The Conditional

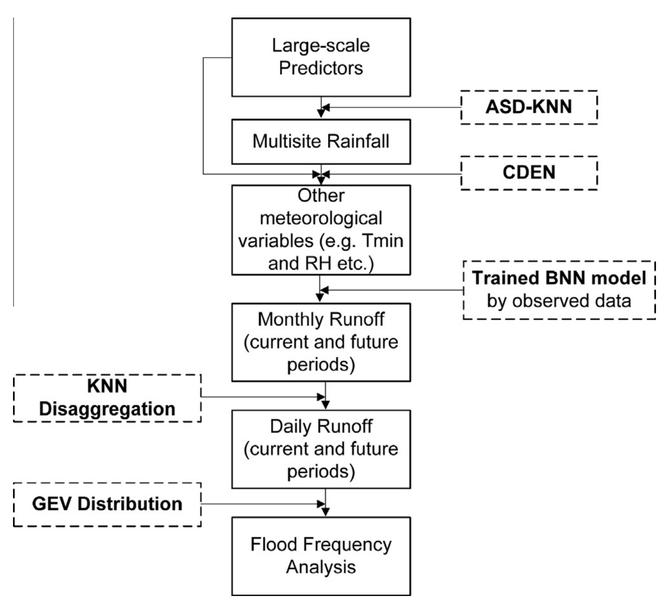

Fig. 1. System diagram of the integrated statistical and data-driven (ISD) framework.