Journal of Hydrology 562 (2018) 1–16

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Influence of rainfall data scarcity on non-point source pollution prediction: Implications for physically based models

Lei Chen[a], Jiajia Xu[a], Guobo Wang[b], Hongbin Liu[c], Limei Zhai[c], Shuang Li[a], Cheng Sun[a], Zhenyao Shen[a,*]

[a] *State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, PR China*
[b] *Waterborne Transport Research Institute, Ministry of Transport of the People's Republic of China, Beijing 100088, PR China*
[c] *Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Key Laboratory of Non-Point Source Pollution Control, Ministry of Agriculture, Beijing 100081, PR China*

A R T I C L E I N F O

This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of Hazi Mohammad Azamathulla, Associate Editor

Keywords:
Information entropy
Non-point source pollution
Rainfall data
Scarcity
Soil and Water Assessment Tool
Three Gorges Reservoir Region

A B S T R A C T

Hydrological and non-point source pollution (H/NPS) predictions in ungagged basins have become the key problem for watershed studies, especially for those large-scale catchments. However, few studies have explored the comprehensive impacts of rainfall data scarcity on H/NPS predictions. This study focused on: 1) the effects of rainfall spatial scarcity (by removing 11%–67% of stations based on their locations) on the H/NPS results; and 2) the impacts of rainfall temporal scarcity (10%–60% data scarcity in time series); and 3) the development of a new evaluation method that incorporates information entropy. A case study was undertaken using the Soil and Water Assessment Tool (SWAT) in a typical watershed in China. The results of this study highlighted the importance of critical-site rainfall stations that often showed greater influences and cross-tributary impacts on the H/NPS simulations. Higher missing rates above a certain threshold as well as missing locations during the wet periods resulted in poorer simulation results. Compared to traditional indicators, information entropy could serve as a good substitute because it reflects the distribution of spatial variability and the development of temporal heterogeneity. This paper reports important implications for the application of Distributed Hydrological Models and Semi-distributed Hydrological Models, as well as for the optimal design of rainfall gauges among large basins.

1. Introduction

Non-point source (NPS) pollution has been a key threat to water quality for decades, and hydrological and non-point source (H/NPS) models are the main tools used to quantify NPS pollution (Andréassian et al., 2001; Bera and Borah, 2003; Chaubey et al., 1999; Drecht et al., 2003). Typically, H/NPS models can be divided into Lumped Hydrologic Models (LHMs), Distributed Hydrological Models (DHMs), and Semi-distributed Hydrological Model (SDHMs). The LHMs regard the watershed as a whole object and cannot reflect the spatial heterogeneity of the actual process inside the watershed (Hrachowitz and Clark, 2017). Conversely, DHMs/SDHMs divide watersheds into smaller spatial units, while pollutants are calculated from each separate unit and are then summed at the watershed outlet. As one special kind of DHMs, the SDHMs divided the entire basin into sub-watersheds first and then into a number of hydrological response units (HRUs) or other computational units depending on slope, soil type and land use instead of rectangular grids (with uniform size) (Hrachowitz et al., 2016; Viviroli et al., 2009). In this sense, spatial variations in climate, underlying surfaces and related hydrological elements could be considered, and spatial data with higher accuracy that are typically derived via remote sensing (RS) and geographic information system (GIS) technologies are required (Bieger et al., 2014; Wang et al., 2016). The commonly used DHMs/SDHMs include the Soil and Water Assessment Tool (SWAT) model, the Institute of Hydrology Distributed Model (IHDM) and TOPMODEL. Among these models, the SWAT model has become one of the most widely used tools in describing temporal and spatial variations in H/NPS cycles, especially for large-scale watersheds due to their greater heterogeneities (Amatya et al., 2011).

Rainfall data are regarded as the most important inputs for DHMs/SDHMs because they act as the driving force of runoff generation and pollutant transportation (Lobligeois et al., 2014; Kashani et al., 2016; Sun et al., 2017). Typically, rainfall data could be obtained using both rainfall station and radar product (Kashani et al., 2016; Pereiracardenal et al., 2011). The application of the radar product has become more widespread as radar technique can reflect the spatial and temporal

* Corresponding author.
E-mail addresses: chenlei1982bnu@bnu.edu.cn, zyshen@bnu.edu.cn (Z. Shen).

https://doi.org/10.1016/j.jhydrol.2018.04.044
Received 4 February 2018; Received in revised form 15 April 2018; Accepted 17 April 2018
Available online 30 April 2018

rainfall variability, especially for those remote regions (Biggs and Atkinson, 2015). However, the radar rainfall product is sometime questioned as its resolutions are generally based on quantitative rainfall estimates (QPEs) from weather radar networks (Harrison et al., 2009). Several major disadvantages of QPEs, such as low spatial resolution, low level forecast for torrential rain and rough prediction of spatial-temporal structure of heavy rainfall, make the radar data cannot meet the requirements of H/NPS models (Cluckie et al., 2004). Thus, there is a growing need for combined (synergic) use of radar measurements along with the rainfall stations, and the rain gauge is still the fundamental source used by DHMs/SDHMs to determine rainfall variability.

Previous studies have indicated that rainfall has irregular changes due to the varying natural conditions and shows strong spatial-temporal heterogeneities among large-scale watersheds, which would impact simulated results of the DHMs/SDHMs (Bardossy and Plate, 1992; Shen et al., 2012). Long time series of rainfall data is often required for comprehensive H/NPS evaluations by covering high flow, normal flow and low flow periods to the greatest extent (Kuangyao et al., 2000; Troin et al., 2012). However, automatic weather stations are susceptible because certain time series may be lost in part due to direct and indirect damage caused by lightning electromagnetic pulses. Electromagnetic interference and human operational errors are also destructive factors for automatic stations and can result in random and continuous absences of rainfall data. As a result, damage to automatic stations and other equipment errors due to various human and natural factors can cause losses in rainfall data series (Shen et al., 2015a; Wambura et al., 2017). On the other hand, previous study has indicated spatial rainfall variability is an important source of NPS simulation uncertainty; thus, users must focus on the optimal design of rainfall stations (Shen et al., 2012). However, accurate descriptions of spatial rainfall variability cannot be obtained due to the paucity of existing observation sites, which is caused by economic and technique conditions (i.e., terrain). The spatial and temporal resolutions of rainfall data are constrained and interrelated and will affect the quantification of H/NPS (Meusburger et al., 2012; Michaelides et al., 2009). Thus, rainfall data scarcity indeed exists and has become the key barrier to H/NPS prediction.

H/NPS predictions in ungauged basins (PUB) have become a hot topic for hydrological researchers because data scarcity is still a worldwide problem. The International Association of Hydrological Sciences (IAHS) began the PUB programme as its first key research programme in the 21st century, which is called as ‘PUB, 2003–2012: Shaping an exciting future for the hydrological sciences.’ (Hrachowitz et al., 2013). For those DHMs/SDHMs, scarce scenarios of rainfall data could be divided into two types: spatial series and time series. Satellite remote sensing play an important role in the PUB (Lakshmi, 2004), but it cannot solve the problem of high spatial resolution and frequent time duplication (Lakshmi, 2016). Studies have also focused on the effect of different spatial interpolation methods and their combinations on rainfall estimation during different rainfall periods (Ali et al., 2005; Cheng et al., 2017; Swain and Patra, 2017). Although these studies have shown that spatial rainfall variability could be obtained by interpolating rainfall data of each station by the methods such as Thiessen polygons, average method and centroid method (Cho et al., 2009), there are few studies focused on their performances during data scarcity scenarios specifically. Weather generators have been developed for DHMs/SDHMs to cope with data scarcity in time series, but the impacts of data scarcity scenarios on their interpolation abilities are not clear yet (Chen et al., 2017a; Ruan et al., 2016). In that regard, quantifying the impacts of the temporal and spatial scarcity of rainfall data is indeed crucial because those data provide the basic inputs for H/NPS models.

Therefore, we focused on the influences of rainfall data scarcity on total phosphorus (TP) prediction in a NPS-dominant catchment by running and comparing different rainfall scarcity scenarios. The following tasks have been performed by: 1) quantifying the effects of rainfall spatial scarcity (station number and location) on H/TP simulations, and 2) exploring the impacts of rainfall temporal scarcity (data scarcity in time series) on H/TP predictions, and 3) developing the information entropy as a new method to replace traditional indicators for data scarcity evaluation. The study is carried out for Daning watershed, China.

2. Methods and materials

2.1. Study area description and data collection

2.1.1. Study area description

The Daning River watershed is located in Wushan county and Wuxi county, the Municipality of Chongqing, China. As an important tributary in the Three Gorges Reservoir area, the drainage area of the Daning River watershed is approximately 2422 km^2 and consists of four major tributaries, including the Xixi River, the Houxi River, the Boyang River and the Dongxi River. The mean rainfall is 1030–1950 mm per year and the annual temperature of the watershed is approximately 18.4 °C. The elevation of the Daning River watershed ranges from 200 m to 2605 m, while the land use types in this area are primarily comprised of forest lands (61.85%), croplands (24.90%) and grasslands (12.48%). This area also consists of seven major soil types: yellow soils (46.05%), yellow-brown soils (25.79%), limestone soils (18.19%), brown soils (6.28%), purple soils (1.98%), paddy soil (1.52%) and alluvial soils (less than 0.2%). The location of the watershed in the Three Gorges Reservoir area and other information are shown in Fig. 1. Based on historical record, the Daning River suffers serious NPS pollution, and phosphorus (P) is the limiting nutrient causing eutrophication in the Three Gorges Reservoir Region (Shen et al., 2012, 2015b). In this area, excessive use of phosphate fertilizer are used on slope cropland and the rainfall would cause P being carrying into the receiving body of water through rainfall erosion and runoff transport from non-specific sites in form of NPS pollution. Thus, TP was selected as a representative of NPS pollutants and water quality in this region.

2.1.2. Data collection

For a comprehensive evaluation, detailed data that are available in this region are collected and compiled as follows:

- Daily rainfall data at nine rainfall stations inside the watershed and four stations at sites approximately 30 km outside the watershed boundary from 1998 to 2008 were obtained from the Meteorological Bureau of Wuxi County and China National Meteorological Administration.
- Other meteorological data from 2000 to 2008, such as daily maximum and minimum air temperature, relative humidity, wind speed and sunlight radiation, were collected at the Meteorological Bureau of Wuxi County.
- The Digital Elevation Model (DEM), which had a resolution of 1:250,000, was digitized from raw data provided by the National Fundamental Geographic Information Center of China.
- The land use map, which had a resolution of 1:100,000, was obtained from the Resources and Environment science data Center of the Chinese Sciences Academy.
- The soil type map, which had a resolution of 1:50,000, was obtained from the Agricultural Science Committee of Wuxi city. The soil physical properties, including soil density, saturated and unsaturated soil hydraulic conductivity, and field capacity, were acquired from the Institute of Soil Science in Nanjing. The soil chemical properties were obtained from the Soil Database of China.
- The crop management measures were obtained from field investigations with local farmers.
- The measured daily flow data and monthly water quality data are mainly obtained from the Yangtze River Basin Water Conservancy Commission and the Wuxi County Environmental Protection