Journal of Hydrology 550 (2017) 616–634

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Evaluation of long-term trends in extreme precipitation: Implications of in-filled historical data use for analysis

Ramesh S.V. Teegavarapu [a,*], Anurag Nayak [b]

[a] Department of Civil, Environmental and Geomatics Engineering, Florida Atlantic University, Boca Raton, FL 33431, United States
[b] Sutron Corporation, West Palm Beach, FL, United States

ARTICLE INFO

Article history:
Received 2 December 2015
Received in revised form 22 March 2017
Accepted 15 May 2017
Available online 19 May 2017
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Bruno Merz, Associate Editor

Keywords:
Extreme precipitation
Trend analysis
Missing data
Infilling methods
Atlantic Multidecadal Oscillation (AMO)
South Florida

ABSTRACT

This study focuses on the assessment of biases from infilling missing precipitation data on the detection of long-term change using parametric and non-parametric statistical techniques. Long-term historical precipitation data available for almost 100 years at 53 rain gages in south Florida, USA, with gages having varying lengths of missing data are used for the study. Precipitation data with gaps and time series with spatial interpolated data are analyzed. Chronologically complete datasets are often used in climate variability studies by analyzing data in multiple temporal windows. The temporal windows selected in this work coincide with Atlantic multi-decadal oscillation (AMO) cool and warm phases that strongly influence precipitation extremes and characteristics in the study region. Selection of these windows has helped in evaluating the extremes derived based on infilled and unfilled data. The frequency of occurrence of precipitation extremes over a pre-specified threshold is also analyzed. Results indicate that infilled precipitation data introduce large biases in the statistical trends and over and under-estimate low and high extremes respectively. Evaluation of three extreme precipitation indices (i.e. Rx1day, R25mm and R50mm) indicates that bias increases with increase in amount of missing data. Nonparametric hypothesis tests indicate that statistical distributions of data of infilled and unfilled data are different when the data infilled is more than 5% of the entire data. Infilled data also introduced high variability in precipitation extremes in AMO cool and warm phases along with the changes in the frequency of occurrence of extreme events over a threshold.

1. Introduction

Long-term data on precipitation occurrence, intensity, amount, and spatial and temporal distribution are vital for the design of hydrologic structures (such as dams, culverts, detention basins, etc.), water supply and water quality modeling, and other hydrologic studies. Assessment of long-term trends in extreme precipitation data is critical for hydrologic design. An understanding of changing extremes is also essential from climate variability and change perspectives and requires an extensive evaluation of trends in entire or temporal slices of historical precipitation data series. Analysis of long-term extreme precipitation data is essential to address issues related to expected and observed changes in frequency and intensity of extreme events in response to warming climate conditions. Several research studies have reported increases in extreme precipitation and flood events over the contiguous United States during the last century (e.g. Changnon and Kunkel, 1995; Karl and Knight, 1998; Kunkel et al., 1999; Groisman et al., 2001). Recent studies by Teegavarapu et al. (2013) and Goly and Teegavarapu (2014) have reported comprehensive evaluation of spatial and temporal variability of precipitation extremes linked with coupled oceanic-atmospheric oscillations in the state of Florida, USA. Their studies point to uniform and non-uniform spatial variations of precipitation extremes at different temporal resolutions in their case study region along with possible influences of regional hydroclimatology on the spatial extent of influences of oscillations on the precipitation characteristics and extremes. Furthermore, strong association of sea surface temperature (SST) anomalies with regional and global climate has also been well documented in several studies (e.g. McCabe et al., 2004; Rogers and Coleman, 2003). The relationship between SSTs and variability in precipitation over the United States and other regions of the world has been discussed in several research studies (Enfield et al., 2001; Goodrich and Ellis, 2008; Lachniet et al., 2004). Teleconnections described as associations between climatic variations (i.e., anomalies) at distant locations

* Corresponding author.
E-mail addresses: rteegava@fau.edu (R.S.V. Teegavarapu), nayakanurag@gmail.com (A. Nayak).

http://dx.doi.org/10.1016/j.jhydrol.2017.05.030
0022-1694/

influence regional precipitation patterns in the study region of interest reported in this paper. Three major teleconnections the Atlantic Multidecadal Oscillation (AMO), El Niño Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) heavily influence precipitation extremes in Florida. A comprehensive study evaluating links between warm and cool phases of AMO and ENSO to precipitation characteristics and extremes has been reported recently by Teegavarapu et al. (2013) and Goly and Teegavarapu (2014). Spatially non-uniform and uniform influences of AMO and ENSO respectively on precipitation extremes and characteristics were noted in the state of Florida. Enfield et al. (2001) demonstrated that the cool (warm) AMO phases are directly related to the above (below) normal rainfall over most of the United States except a few regions in the southeast. Strong correlation of ENSO phases with the increase (decrease) in precipitation and snowfall over United States and Canada has been documented in several research findings (e.g. Groisman and Easterling, 1994; Groisman et al., 1994; Latif and Barnett, 1994).

In a recent study Teegavarapu et al. (2013) evaluated the influence of two different phases (warm and cool) of AMO on precipitation extremes in Florida. They concluded that magnitudes of precipitation extremes from one specific phase (i.e., warm phase) are higher than those obtained from the two phases combined. Therefore, it is important that the analysis period for evaluation of precipitation extremes should include both warm and cold phases of AMO. Since long-term variations in precipitation data are of interest in this study, understanding the differences in extremes values derived from data with gaps and infilled data is critical. Also, it is essential to investigate if inferences made about influences of oscillations on these extremes are different when these data sets with and without gaps are used in climate variability studies.

Availability and quality of long-term gap-free precipitation data at temporal resolutions of day or less required to perform such analyses are often limited. Long-term precipitation datasets generally contain gaps due to instrument malfunctioning, inability of operator to collect data, discontinuity and a variety of other reasons. Statistical analyses of long-term precipitation data require filling of these data gaps with appropriate values to obtain a serially complete dataset. Data filling techniques are normally selected based on location and desired accuracy required for the hydrologic analyses. Missing data filling methods are also based on consideration given to the nature of missing data defined by one of the three conditions (Rubin, 1987, 1996; Little and Rubin, 2002; Graham, 2012): (1) missing at random (MAR); (2) missing completely at random (MCAR) and (3) missing not at random (MNAR). The conditions MAR and MCAR generally indicate that the probability of having a missing observation does not depend on observed values or unobserved values. MNAR is opposite to the two conditions (i.e., MCAR and MAR). Mapping the mechanisms of for missing data is generally difficult. One of the most straightforward definitions for MCAR is "... missingness does not depend on the values of the data, missing or observed ..." (Little and Rubin, 2002). This definition is used in the current study.

Spatial interpolation techniques (such as inverse distance weighted, non-linear, deterministic, stochastic interpolation (e.g. Kriging) and data-driven methods (regression and time series analyses) are generally used for estimation of missing precipitation data (Karl and Knight, 1998; Brunetti et al., 2001; Teegavarapu and Chandramouli, 2005; Teegavarapu, 2009; Teegavarapu, 2013, 2012a). Other available methods include normal-ratio and inverse distance weighting methods for estimation of missing data (ASCE, 1996; Teegavarapu and Chandramouli, 2005). Kriging in various forms has been used to estimate missing precipitation data as well as to interpolate precipitation from point measurements (Dingman, 2002; Vieux, 2001; Ashraf et al., 1997; Teegavarapu, 2007). Statistical distribution-based missing data estimation methods (e.g. Brunetti et al., 2001) are also available. Karl and Knight (1998) and Brunetti et al. (2001) report the use of Gamma distribution for estimating missing daily precipitation values. Woolhiser and Pegram (1979), Wilson et al. (1992) and Hanson et al. (1994) used a bivariate exponential distribution to fill the missing precipitation data. Several problems exist with the application of spatial interpolation methods and some of the major ones are: "tent pole effect" (Vieux, 2001; Teegavarapu, 2012b) that refers to existence of larger values nearer to the control points after estimation, uncertainty associated with selection of neighbors, use of inappropriate variograms in kriging, negative estimates while using artificial neural network (ANN) methods and thin splines and trend surface models, arbitrary values assigned to sill and nugget parameters in kriging resulting in artifacts in interpolation, observation value-insensitive variance estimates, and the computational burden to interpolate the surfaces.

Spatial interpolation or temporal interpolation methods can be used for infilling missing data in precipitation time series. The former method uses observations available at different sites in a region for infilling the data at a site with missing data (i.e. base site), while the latter method employs only data from the site (base site) itself to infill data. Success of spatial interpolation is attributed to existence of strong spatial correlation between any site and base site. Temporal interpolation depends on existence of serial autocorrelation in precipitation time series for developing linear or nonlinear interpolation models or autoregressive models. In some cases, spatial interpolation can only be used as temporal interpolation fails due to lack of high serial correlation at several lags in daily precipitation data. Data filling can lead to changes in the probability distribution of data at a site and introduce significant biases when event-based analysis (such as daily or hourly extreme precipitation analysis) is performed. At temporal resolutions of a day or less, spatial interpolation alters the probability distribution of data, changes the autocorrelation structure and dry and wet spell transitions (Teegavarapu, 2014). Many research efforts have focused on analyzing the trends in precipitation data, especially extremes, but biases introduced in these trends due to data infilling techniques are rarely investigated. Previous study by Teegavarapu (2014) used only a fixed length of missing data and analyzed the impact of infilling on precipitation characteristics and extremes. The current study employs: (1) varying lengths of missing data for evaluation of statistically significant trends; (2) resampling methods to quantify the impacts of infilling on evaluation of extremes and (3) nonparametric hypothesis tests for assessment for differences in distributions of filled and unfilled precipitation datasets for different lengths of gaps. All these tasks were not carried out in the previous study. In this study, long-term precipitation data collected at 53 National Oceanographic and Atmospheric Administration (NOAA) rain gages (shown in Fig. 1) in south Florida, U.S., were analyzed. These data contained several gaps (missing data of varying lengths at different gages) and have been infilled using a spatial interpolation technique (Aly et al., 2009). These serially complete precipitation data are meant to be used in various hydrological and hydraulics modeling studies in the region. In this study, the biases introduced due to infilling of missing data are investigated. Also, the long-term trends in extreme precipitation indices are evaluated at 14 select rain gages for the 1945–2006 analysis period where data infilling was found to be minimal.

The main objectives of the study are: 1) evaluation of bias in precipitation extremes due to infilling of data gaps; 2) evaluation of changes in long-term trends in extreme precipitation indices due to infilling; 3) evaluation of probability distributions of infilled