Journal of Hydrology 509 (2014) 55–65

Contents lists available at ScienceDirect

Journal of Hydrology

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

Multisite rainfall downscaling and disaggregation in a tropical urban area

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

[a] School of Civil & Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b] Earth Observatory of Singapore (EOS), Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 24 June 2013
Received in revised form 25 October 2013
Accepted 16 November 2013
Available online 22 November 2013
This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Magdeline Laba, Associate Editor

Keywords:
Downscaling
Disaggregation
GLM
KNN
HYETOS
MuDRain

SUMMARY

A systematic downscaling-disaggregation study was conducted over Singapore Island, with an aim to generate high spatial and temporal resolution rainfall data under future climate-change conditions. The study consisted of two major components. The first part was to perform an inter-comparison of various alternatives of downscaling and disaggregation methods based on observed data. This included (i) single-site generalized linear model (GLM) plus K-nearest neighbor (KNN) (S-G-K) vs. multisite GLM (M-G) for spatial downscaling, (ii) HYETOS vs. KNN for single-site disaggregation, and (iii) KNN vs. MuDRain (Multivariate Rainfall Disaggregation tool) for multisite disaggregation. The results revealed that, for multisite downscaling, M-G performs better than S-G-K in covering the observed data with a lower RMSE value; for single-site disaggregation, KNN could better keep the basic statistics (i.e. standard deviation, lag-1 autocorrelation and probability of wet hour) than HYETOS; for multisite disaggregation, MuDRain outperformed KNN in fitting interstation correlations. In the second part of the study, an integrated downscaling-disaggregation framework based on M-G, KNN, and MuDRain was used to generate hourly rainfall at multiple sites. The results indicated that the downscaled and disaggregated rainfall data based on multiple ensembles from HadCM3 for the period from 1980 to 2010 could well cover the observed mean rainfall amount and extreme data, and also reasonably keep the spatial correlations both at daily and hourly timescales. The framework was also used to project future rainfall conditions under HadCM3 SRES A2 and B2 scenarios. It was indicated that the annual rainfall amount could reduce up to 5% at the end of this century, but the rainfall of wet season and extreme hourly rainfall could notably increase.

1. Introduction

High-resolution spatial and temporal rainfall data is essential for studies of water resources management, hydrological modeling, and flood risk assessment. This is especially true for tropical urban areas where highly complex rainfall patterns exist (Abustan et al., 2008). The previous studies on climate variables and their implications to runoffs have highlighted the necessity to have input data at short timescales for many hydrological models (Mezghani and Hingray, 2009). However, high-resolution data is limited at many regions due to restrictions of cost, technical capability and physical geographic condition. It is also challenging to conduct high-resolution impact studies for hydrological systems under climate change, due to the coarse resolution of general circulation models (GCMs). Using statistical methods (such as spatial downscaling and temporal disaggregation methods) to generate high-resolution rainfall data has demonstrated a viable alternative and the number of the related studies has increased dramatically in the past years.

The fundamental concept of statistical downscaling is to build a linkage between the variables of GCMs at a large scale (predictors) and local observed weather information (predictands) (Fowler et al., 2007). The widespread used downscaling models could be classified into three types: (i) linear regression models, such as statistical downscaling model (SDSM) (Wilby et al., 2002), generalized linear model (GLM) (Chandler and Wheater, 2002), and automated statistical downscaling tool (ASD) (Hessami et al., 2008); (ii) nonlinear regression models, such as artificial neural network (ANN) (Zorita and von Storch, 1999) and support vector machine (SVM) (Tripathi et al., 2006); (iii) weather generators, such as Long Ashton research station-weather generator (LARS-WG) (Racsko et al., 1991), 'Richardson' type weather generator (WGEN) (Wilks, 1992), and agriculture and agri-food Canada-weather generator (AAFC-WG) (Hayhoe, 2000). Among many alternatives, GLM is an effective stochastic rainfall model based on linear regression, and

* Corresponding author at: School of Civil & Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore. Tel.: +65 67905288; fax: +65 67921650.
E-mail address: xsqin@ntu.edu.sg (X.S. Qin).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.11.027

has proved to be advantageous in addressing issues of spatial correlation, site effect, and seasonal variations, etc. Chandler and Wheater (2002) applied GLM to downscale atmospheric predictors at western Ireland, where logistic regression and gamma distribution were used for occurrence and amount modeling, respectively. Yang et al. (2005) employed GLM to generate daily rainfall at southern England, and showed that the model could reproduce properties at a scale ranging over 2000 km^2. Tisseuil et al. (2011) used GLM, generalized additive model (GAM), aggregated boosted trees (ABT), and multi-layer perceptron neural networks (ANN) to downscale precipitation and evaporation at southwest France. The results showed that the three non-linear models had a better performance than GLM for modeling fortnightly flow percentiles. Beuchat et al. (2012) applied GLM with weighting schemes for downscaling rainfall at 27 sites covering Switzerland. The results showed that the downscaled rainfall exhibited a spatially coherent pattern at seasonal timescale, although spatial independence was assumed by the GLM method.

Many studies also focused on generation of rainfall at a finer timescale using different disaggregation methods. The major types include stochastic point process models (Rodriguez-Iturbe et al., 1987, 1988; Khaliq and Cunnane, 1996; Heneker et al., 2001; Debele et al., 2007; Engida and Esteves, 2011), non-parametric resampling models (Prairie et al., 2007; Nowak et al., 2010; Kalra and Ahmad, 2011) and others (Gyasi-Agyei, 2005; Gyasi-Agyei, 2011; Beuchat et al., 2011). Among these models, HYETOS and K-nearest neighbors (KNN) were widely used. Koutsoyiannis and Onof (2001) developed a hybrid model based on the Bartlett-Lewis rectangular pulses model, called HYETOS. It added an adjustment procedure to assure the sum of disaggregated hourly data be consistent with the given daily data. Debele et al. (2007) applied HYETOS to disaggregate daily rainfall to hourly ones at Cedar Creek watershed in US. Prairie et al. (2007) explored a stochastic nonparametric method, KNN, for spatial-temporal disaggregation of stream flows, and indicated that the KNN method could guarantee the simulation of statistical properties in the original space (historical record). Kalra and Ahmad (2011) applied KNN nonparametric method to generate seasonal precipitation by disaggregating annual precipitation, and the study results indicated that the KNN method performed better than the first-order periodic autoregressive parametric approach, and the seasonal precipitation reproduced on winter and spring seasons was more satisfactory. These studies focused on single-site disaggregation. For multiple sites, the cross-correlation becomes an important factor to be considered. Some studies attempted development of stochastic weather generators for multi-site rainfall generations (Wilks, 1998; Burton et al., 2008; Jennings et al., 2010), but most of them were not able to deal with disaggregation at the same time. As a viable attempt, Koutsoyiannis et al. (2003) developed a method, called MuDRain, by combining a simplified multivariate rainfall model and transformation model to disaggregate daily rainfall to hourly ones at multiple sites. In the study of Debele et al. (2007), MuDRain model was applied and showed an outperformed result for reproduction of expected statistical properties (e.g. average hourly rainfall, standard deviation, probability of wet hour and skewness) with small RMSE values, especially for the reproduction of peak value and temporal distribution; more importantly, the inter-site cross-correlation could be captured very well.

Based on the above review, it is recognized that many hydrological applications require a full spatial distribution of rainfall at finer timescale. This is especially true for climate change impact studies, where the global circulation models could only offer projections at coarse spatial and temporal resolutions. Hence, integrated downscaling and disaggregation effort is necessary as it provides rainfall data with both high spatial and temporal resolutions to meet the requirement of hydrological modeling. There are relatively few studies in such an area. Segond et al. (2006) proposed a combined spatial-temporal downscaling and disaggregation approach using GLM, HYETOS and an artificial profile multisite transformation method. In this approach, the daily data was generated by GLM for multiple sites; HYETOS was used for disaggregating daily data to hourly ones at the master station which contained a historical hourly record; then, the disaggregated hourly data pattern was projected to other sites (i.e. satellite stations) using the artificial profile method. By comparing with the observed values, the simulated rainfall preserved the desired statistical properties (e.g. mean, standard deviation, autocorrelation, probability of dry day/hour, skewness and cross-correlations) with relatively low errors (e.g. between 0.3% and 18%), and the generated envelope could well cover the observed data. Mezghani and Hingray (2009) developed another combined downscaling-disaggregation approach for both temperature and rainfall over the Upper Rhone River basin in the Swiss Alps. GLM was used for downscaling mean areal weather variables (including total precipitation, rainfall and temperature) from GCM model, and KNN was used for disaggregating them to sub-daily and sub-regional scales. The study results showed a good performance of the proposed method in generating statistical relationship, including spatial cross-correlations.

Generally, the integration of spatial downscaling and temporal disaggregation could offer high-resolution rainfall data projected from GCM scenarios, and has great potential to help examine the impact of climate change on rainfall patterns and hydrological systems. From reviewing the recent research works, it turns out that such studies are relatively limited. Essentially, there is a lack of an inter-comparison study that could show the advantages or limitations of various options of single-site or multisite rainfall downscaling and disaggregation techniques that could keep the key statistics at different time scales, particularly in connection with the output from a downscaling model. In addition, most of the previous studies focused on a relatively larger scale watershed or basins. There are limited efforts on integrated downscaling and temporal disaggregation for the urban areas at Southeast Asia, which is characterized by tropical climate with rainfalls showing high temporal-spatial variations.

Therefore, the objective of this research work is to conduct a systematic rainfall downscaling-disaggregation study at a tropical urban area (i.e. Singapore Island). An inter-comparison study will be performed first to evaluate various options in implementing single and multiple site downscaling and disaggregation, based on statistical indicators (at daily and hourly scales) and observed data. Downscaling will essentially be based on GLM model as it has already been proved as an advantageous tool in keeping many key daily statistics of rainfall (Yang et al., 2005). Options of KNN, HYETOS, and MuDRain will be tested for temporal disaggregation. Based on the comparison result, the deemed best option of downscaling-disaggregation framework will be used for projecting high-resolution rainfall patterns under future climate conditions for the Singapore Island.

2. Study area and data

Singapore, with an area of about 723 km^2, locates at the equator pluvial region. Most of the surface elevation over the island is below 15 m, and the highest point is Bukit Timah hill which has a height of 165 m at the central region. The small hill leads to a 'rain shadow' phenomenon (Whiteman, 2000) that induces slight disparities of weather distribution on the western and eastern sides of the island (e.g. the western side of Singapore is wetter than the eastern one). Singapore has a rich precipitation, with an average annual rainfall amount being more than 2400 mm. The highest record of daily rainfall was near 520 mm which happened at the