Environmental Modelling & Software 67 (2015) 57–71

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka

Cameron C.F. Plouffe [a, *], Colin Robertson [a], Lalith Chandrapala [b]

[a] Department of Geography & Environmental Studies, Wilfrid Laurier University, 75 University Ave West, Waterloo, ON, N2L 3C5, Canada
[b] Department of Meteorology, 383 Bauddhaloka Mawatha, Colombo 7, Sri Lanka

A R T I C L E I N F O

Article history:
Received 6 November 2014
Received in revised form
12 January 2015
Accepted 19 January 2015
Available online

Keywords:
Spatial interpolation
Rainfall prediction
Kriging
Raster comparison
Sri Lanka

A B S T R A C T

Interpolating climatic variables such as rainfall is challenging due to the highly variable nature of meteorological processes, the effects of terrain and geography, and the difficulty in establishing a representative network of stations. While interpolation models are being adapted to include these effects, often the rainfall data contain significant gaps in coverage. In this paper, we evaluated rainfall data from an agro-ecological monitoring network for producing maps of total monthly rainfall in Sri Lanka. We compared four spatial interpolation techniques: inverse distance weighting, thin-plate splines, ordinary kriging, and Bayesian kriging. Error metrics were used to validate interpolations against independent data. Satellite data were used to assess the spatial pattern of rainfall. Results indicated that Bayesian kriging and splines performed best in low and high rainfall, respectively. Rainfall maps generated from the agro-ecological network were found to have accuracies consistent with previous studies in Sri Lanka.

1. Introduction

Ecological forecast models that rely on climatic variables are increasingly used in a variety of contexts. For example, detailed climate data are necessary when modelling outbreak patterns of emerging infectious diseases (Briët et al., 2008; Robertson et al., 2012). While data are increasingly becoming available due to the advent of smaller and cheaper environmental sensors, most climate data – specifically, precipitation data – are still collected by a network of geographically dispersed weather stations. This leads to data that contain considerable gaps in coverage of areas where stations are more isolated. However, additional data sources such as citizen sensors (Goodchild, 2007), unofficial and/or semi-official networks of rain gauges (Wickramaarachchi et al., 2013), and satellite-derived data products (Kummerow et al., 1998) may be used to augment estimates from ground-based stations. In order to leverage these auxiliary sources of data, new approaches are required to integrate data from multiple sources, and to help evaluate the best performing models (Bennett et al., 2013). In this paper, we investigate the integration of additional sources of data for comparison of rainfall interpolation methods in Sri Lanka. Firstly, we aim to evaluate an unofficial network of rain gauges across Sri Lanka by comparing different interpolations against official meteorological station recordings. Secondly, as part of our validation, we examine spatial patterns in predicted rainfall in relation to satellite-derived estimates of rainfall.

Spatial interpolation techniques are widely used to estimate seamless spatial coverage of rainfall over large areas, yet there is little consensus on the optimal interpolator for rainfall, especially where spatial rainfall pattern is highly variable (Dirks et al., 1998; Price et al., 2000; Vicente-Serrano et al., 2003). Table 1 displays a summary of several different studies evaluating interpolation methods applied for rainfall prediction in different settings. Previous studies have come to different conclusions regarding the most effective techniques for spatial interpolation of rainfall data, and more generally, measuring performance for any given

Abbreviations: IDW, Inverse Distance Weighting; MAE, Mean Absolute Error; MdPE, Median Percent Error; SE, Statistical Error; SRMSE, Standardized Root Mean Square Error; SSIM, Structural Similarity Index; S, Structure component of SSIM; TRMM, Tropical Rainfall Measuring Mission.

* Corresponding author.
E-mail address: plou7570@mylaurier.ca (C.C.F. Plouffe).

http://dx.doi.org/10.1016/j.envsoft.2015.01.011
1364-8152/

environmental model is intrinsically case-dependent (Bennett et al., 2013). Robson (2014) suggests that opting for the simplest model possible is desirable unless it has been found to be inadequate when compared to more complex models. The literature reveals that accuracy of precipitation interpolation varies greatly by region and temporal scale (Table 1). Interpolation errors are related to measurement error, the density of the station network, topography, and the type of rainfall (Abtew et al., 1993). Tropical and monsoonal environments in particular have proven difficult to characterize with seamless spatial coverages of rainfall (Jayawardene et al., 2005). The amount of rainfall in the tropics is often highly variable in intensity and seasonality (Malhi and Wright, 2004), and interpolating rainfall for these areas can be quite difficult, as climate often dramatically changes over space and time.

The primary differences between the statistical methods used to interpolate rainfall are how they are conceptually formulated and mathematically constructed (Burrough and McDonnell, 1998). Some approaches to spatial interpolation are more effective at predicting certain types of spatial processes, and thus context-specific applications of interpolation methods are common. Comparative studies have been conducted to determine which method of spatial interpolation is best suited for different contexts, but as of yet, no decisive conclusions have been made (Zimmerman et al., 1999). It is important to continue research in this direction in order to gain a better understanding of proper applications of these interpolation techniques.

The tropical country of Sri Lanka is used here as a case study for our exploration of rainfall interpolators. It was hypothesized that one of the geostatistical methods would yield the most accurate results. A review of relevant literature found that kriging is the most effective interpolation method for precipitation data (Jeffrey et al., 2001; Vicente-Serrano et al., 2003; Zimmerman et al., 1999).

1.1. Objectives

The objectives of this research were three-fold. Firstly, we aimed to determine the most effective spatial interpolation methods for rainfall data for application to countrywide environmental modelling in Sri Lanka. Specifically, we required a methodology for estimating seamless spatial coverage of monthly precipitation. While our focus here is Sri Lanka, we aim to add to the literature on interpolation comparisons, with specific emphasis on tropical areas that exhibit large variability in rainfall throughout the year. To investigate this, four different spatial interpolation methods were evaluated: inverse distance weighting (IDW), thin plate smoothing splines, ordinary kriging, and Bayesian kriging. These methods were chosen on the basis that many studies in the past have employed these techniques in rainfall interpolation (Daly et al., 1994; Dirks et al., 1998; Jeffrey et al., 2001; Oke et al., 2009; Vicente Serrano et al., 2003). The results of these comparisons will be used as input for a spatial–temporal model used for online surveillance and forecasting of waterborne infectious disease risk in Sri Lanka. The second objective was evaluate the suitability of a

Table 1
Summary of studies of rainfall interpolation. Studies are sorted by author, process, the interpolation methods being employed, and the overall findings of each study.

Study	Process	Interpolation methods employed	Findings
Hutchinson (1995)	Interpolated annual rainfall for a region of south eastern Australia	- Thin plate smoothing splines	- Splines required no prior estimation of the spatial auto covariance structure, which could prove beneficial when the data set being used could contain errors distributed across entire spatial network of observation stations
Dirks et al. (1998)	Interpolated rainfall obtained for Norfolk Island, of the coast of Australia	- Areal-mean - IDW - Kriging - Thiessen polygons	- All methods found to perform at a similar level - Thiessen polygons produced most unrealistic results, due to discrete rainfall boundaries - IDW deemed most appropriate method, due to accurate interpolations produced, and low performance requirements
Price et al. (2000)	Interpolated monthly mean climate data for study sites in British Columbia/Alberta, and Ontario/Quebec, Canada.	- ANUSPLIN, a software based on thin plate smoothing splines - GIDS, a regression-based model	- Extreme outliers that exceeded 100% difference between observed and predicted values present for precipitation interpolations - ANUSPLIN produced slightly more accurate predictions, as it could more easily account for changes in elevation - Regions with sparse data occasionally exhibited negative precipitation values being predicted
Vicente-Serrano et al. (2003)	Analysed validity of precipitation and temperature maps of the Ebro Valley in northeast Spain	- Empirical regression models - IDW - Kriging methods - Thiessen polygons - Thin plate smoothing splines - Trend surfaces	- Regression modelling and kriging methods produced the highest correlation between observed and predicted rainfall - Trend surfaces and Thiessen polygons were determined to produce the least accurate predictions according to validation statistics
Oke et al. (2009)	Investigated prediction of rainfall across Australia	- Cokriging - Ordinary kriging - Simple kriging with a locally varying mean	- Prediction errors from all three methods found to be similar (negative errors implying underestimation of gauge rainfall present for all methods) even with the inclusion of satellite-based TRMM rainfall estimates for some methods - Satellite rainfall data potentially improved spatial prediction in areas that were not adequate sampled - Errors were consistently negative in coastal regions, while errors in higher inland areas tended to be positive
Newlands et al. (2010)	Evaluated three interpolation methods for precipitation and temperature across much of the Canadian landmass	- ANUSPLIN, a software based on thin plate smoothing splines - HYBRID inverse-distance/natural-neighbour model - DAYMET Weighted-truncated Gaussian filter	- All models predicted reasonably well, with ANUSPLIN producing the most accurate daily mean precipitation values at a 10 km scale - High error variance for precipitation was exhibited in summer along the coasts and in winter in the Prairies region - Authors recommend employing a Bayesian/mixed models methodology for future climate prediction in Canada