5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

Projection of Future Rainfall in Hong Kong Using Logistic Regression and Generalized Linear Model

Chi-shing Calvin Cheung[1,3], Melissa Anne Hart[2,3], Mervyn R. Peart[3]

Abstract— **This study demonstrates a method to project the future rainfall in a city, Hong Kong, based on future climate scenarios. The procedure consists of using logistic regression (LM) to predict rain occurrence and generalized linear model (GLM) to predict rainfall volume. A perturbation method is then applied to statistically downscale the future projections to a local rainfall study. The future climate projections are given by three General Circulation Models (GCMs): GISS-ER, GFDL-CM2.1 and MRI-CGCM2.3.2. from the Coupled Model Intercomparison Project phase 3 (CMIP3). The logistic regression and generalized linear model were first calibrated using NCEP/NCAR reanalysis and local scale observations data from 1971 – 2000, followed by validation using data from 2001 – 2010. The index of agreement for monthly rain occurrence (days) is > 0.62 and monthly rainfall volume (mm) is > 0.73 for model validation.**

I. Motivation

The objective of this research is to project rain occurrence and rainfall amount under future climate conditions in Hong Kong; this will provide useful information to those investigating these impacts in various sectors and for policy makers to take a proactive approach to adaptation and mitigation measures to climate change.

The major impact of heavy and prolonged rainfall is flooding, especially in lowland areas with a high population and locations with poor drainage systems. Expansion of the urban area and coastal reclamation may also contribute to flooding. Infrastructure may be at risk under extreme rainfall, causing flooding and damage to the drainage system [1]. Rainfall is also related to the occurrence of landslip in Hong Kong [2, 3] with the city experiencing 300 to 400 landslides per year [4].

Historically heavy rain has impacted the agriculture and fishery industries, however most food is now imported to Hong Kong and there is very little local agriculture production. However, heavy rain also affects the surrounding region such as Guangdong and hence still affects food supply to Hong Kong. The impacts of rainfall upon the Hong Kong economy and society have been described by Peart (2004) [5], water supply is a constant challenge in Hong Kong, with a drought in 1963 demonstrating that the storage of water supply is a risk to public health and recreation. These impacts show the importance of understanding the occurrence of rain in Hong Kong, particularly under future climate conditions. Therefore it is important to project the future rainfall and understand the potential impact in Hong Kong.

This study uses a combination of logistic regression to determine rain occurrence, followed by a generalized linear model to predict rainfall amount that provide a better projection in term of the rainfall distribution. [6] demonstrated a statistical downscaling scheme involve logistic regression (LR) and generalized linear model (GLM) to predict the monthly rainfall distribution in the west of Ireland, similar studies also reported the stated method is appropriate to predict monthly precipitation [7, 8, 9].

II. Method

Three types of data were used: 1) historical observations; 2) gridded NCEP/NCAR Reanalysis data and 3) GCM projections of future climate scenarios. All observation data is from the Hong Kong Observatory Headquarters weather station in Tsim Sha Tsui. Data from 1971 to 2000 is used for model calibration. This study selected the simulations output from three GCMs: GFDL-CM2.1, GISS-ER and MRI-CGCM2.3.2 from the Coupled Model Intercomparison Project phase 3 (CMIP3), these models provide the projections of future climate. GCMs utilize many different grid systems, this study is limited to the GCMs grids that cover Hong Kong and the surrounding region.

The procedures to project rain occurrence and rainfall volume for GCM's future scenarios involve a two-step method. First, rain days were identified by a logistic model to determine the probability of rain occurrence from large scale climatic variables. A binary variable was created based on rain occurrence from the observation record: rain

Corresponding author: Chi-shing Calvin Cheung
thisiscalvin@cuhk.edu.hk
[1]IEES, The Chinese University of Hong Kong
[2]The University of New South Wales, Sydney, Australia
[3]Dept of Geography, The University of Hong Kong

= 1, non-rain = 0. Days with rainfall ≥ 0.2 mm were designated rain days. Then, rainfall volume is based on the climate variable from the selected rain days. A generalized linear model (GLM) was used to correlate rainfall volume (from observations) and the large scale climate variables (NNR). The models were first calibrated using observations and NCEP/NCAR reanalysis for the period 1971 to 2000, then each model was validated using data from the period 2001 to 2010. For the future climate scenarios a Rain Occurrence Model is applied to calculate the probability to rain. The identified rain days are then extracted from the dataset to predict rainfall volume using Rainfall Volume Model.

III. EVALUATION

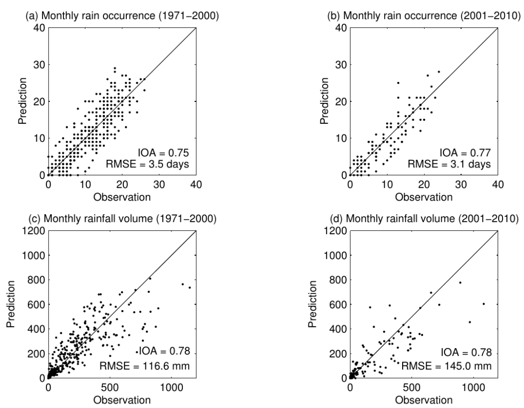

Figure 1 - Scatter plots between observation and prediction (downscaled by NCEP/NCAR reanalysis). (a)-(b) monthly rain day occurrence; (c)-(d) monthly rainfall volume. Calibration period: 1971–2000, validation period: 2001- 2010.

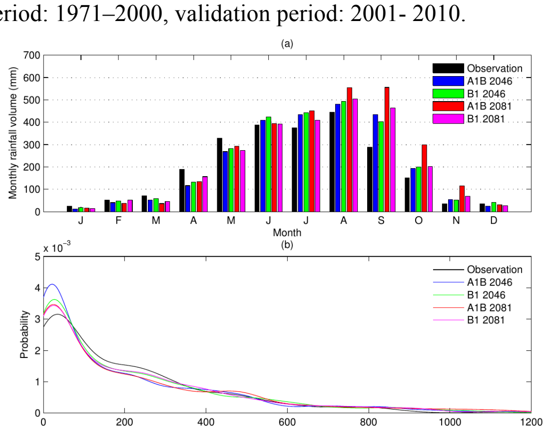

Figure 2 - Multi model ensemble (MME) projection of monthly rainfall volume; (a) Long term average monthly rain occurrence; (b) probability density functions.

Figure 1 shows the performance of the rain occurrence and rainfall volume models. Overall there are good performances in both calibration and validation periods as the index of agreement between prediction and observation reach over 0.75 for monthly rain days and 0.78 for monthly rainfall volume. However, some extreme cases in rainfall volume were under-predicted, indicating that the model worked best below 700mm.

Figure 2 compares the future projections of rainfall volume with the observational record (1971-2000). The future scenarios show less rainfall from February to May and more rainfall from June to November (Figure 2a). Scenario A1B has higher intensity of increased rainfall from June to November compared to scenario B1. From the Probability Density Functions (PDFs) comparison (Figure 2b), there will be higher possibility in extreme low rainfall (below 100 mm) and extreme high rainfall (above 800 mm).

This paper presented results of precipitation downscaling using a two-step approach - logistic regression to determine rain occurrence, followed by a generalized linear model to predict rainfall volume. The logistic model and generalized linear model were calibrated using the observational record (1971 – 2000) and NNR data. It was found that there was good agreement between observations and predictions in terms of monthly rain days and month rainfall volume.

ACKNOWLEDGMENTS

This research was funded through the Hong Kong Research Grants Council (RGC) General Research Fund (GRF) (RGC Ref No. 743908; HKU Project no. HKU 743908H).

REFERENCES

[1] Chan, F., Loh, C., MacDonald, A., Mitchell, G., Adekola, O., 2010. Rich Delta, Costly Flooding. Civic Exchange, Hong Kong.

[2] Brand, E.W., Premchitt, J., Phillipson, H.B., 1984. Relationship between rainfall and landslides in Hong Kong. Presented at the The 4th International Symposium on Landslides, Toronto, pp. 337–384.

[3] Yu, Y.F., Lam, J.S., Siu, C.K., 2003. Interim report on review landslip warning criteria. Geotechnical Engineering Office, Hong Kong Government, Hong Kong SAR.

[4] Parry, S., Ng, K.C., 2010. The assessment of landslide risk from natural slopes in Hong Kong: an engineering geological perspective. Q. J. Eng. Geol. Hydrogeol. 43, 307–320. doi:10.1144/1470-9236/08-012

[5] Peart, M.R., 2004. Water supply and the development of Hong Kong. IAHS-AISH Publ. 23–30.

[6] Chandler, R.E., Wheater, H.S., 2002. Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. Water Resour. Res. 38, 1192. doi:10.1029/2001WR000906 Ambrosino et al., 2010; Buishand et al., 2004; Fealy and Sweeney, 2007.

[7] Ambrosino, C., Chandler, R.E., Todd, M.C., 2010. Southern African Monthly Rainfall Variability: An Analysis Based on Generalized Linear Models. J. Clim. 24, 4600–4617.

[8] Buishand, T.A., Shabalova, M.V., Brandsma, T., 2004. On the Choice of the Temporal Aggregation Level for Statistical Downscaling of Precipitation. J. Clim. 17, 1816–1827.

[9] Fealy, R., Sweeney, J., 2007. Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. Int. J. Climatol. 27, 2083–2094.