Statistical downscaling of future heavy precipitation events for Oahu, Hawai'i using an artificial neural network.

Chase W. Norton, Pao-Shin Chu*, and Thomas A. Schroeder

Department of Meteorology, School of Ocean and Earth Science and Technology

University of Hawaii-Manoa, Honolulu, Hawaii 96822, U.S.A.

Abstract

Using a non-linear artificial neural network (ANN) with a mutlilayer perceptron (MLP) topology, daily extreme precipitation events are statistically downscaled from the ECHAM 5 GCM scenario A2 during the period of 2011-2040. A two-test approach, recommended by the Fourth Assessment Report of the International Panel on Climate Change , is used to determine the most appropriate GCM for downscaling purposes. BCa bootstrap resampling is applied to provide 95% confidence intervals for storm frequency and intensity for all datasets.

The model output suggests an increase in the frequency of heavy rainfall events, but a decrease in the intensity during the next thirty years (2011-2040).

1. Introduction

There is considerable interest in the future climate change for Oahu, Hawai'i because it is the most populous island in the state and future water resource planning and management would have the largest impact in this region. It is located in the Central North Pacific and is greatly influenced by maritime air. Due to external forcings and complex topography from two mountain ranges, precipitation can be highly variable creating regions that range from desert climate to rain forest within only a few miles. Also, the island is vulnerable to periods of heavy precipitation resulting from mid-latitude fronts, Kona storms, upper level disturbances, and tropical cyclones [Kodama and Barnes, 1997].

It is generally agreed upon within the scientific community that the amount of CO_2 in the atmosphere has increased dramatically in the past century as a result of anthropogenic activities [IPCC, 2007]. A global increase in temperature has also been documented in association with CO_2. However, the effect increasing CO_2 has on other variables such as precipitation is not as easily understood.

In an attempt to better understanding wholly the effect of increasing CO_2 on our planet, numerical models known as General Circulation Models (GCMs) were designed and implemented. However, due largely to computational limitations, GCMs have been restricted to very coarse spatial resolution (~250 km). This restricts impact studies, which use the raw data output, to focus on large areas. Island nations and smaller regions require a finer resolution output that is currently only available through methods which downscale the coarse GCM output.

*Corresponding Author
Dr. Pao-Shin Chu, Department of Meteorology, 2525 Correa Road, University of Hawaii, Honolulu, Hawaii 96822; email: chu@hawaii.edu

Statistical downscaling attempts to find an empirical relationship between large-scale atmospheric variables (predictors) and a small-scale variable (predictand). The optimal relationship is found through training and validation of the model. Various statistical downscaling methods are available, but the most common has been linear and non-linear transfer functions.

The current study (Norton et al., 2011) applies a non-linear method known as an artificial neural network (ANN) for downscaling daily precipitation extremes for Oahu, Hawai'i during the time periods 1979-2008 (current climate) and 2011-2040 (future climate). It has been shown that ANNs can be an adequate method when dealing with nonlinear relationships found in meteorology [Hsieh and Tang, 1998].

In order to assess the confidence interval of both empirical and modeled extreme events statistics, a BCa percentile bootstrap resampling method is employed [Efron and Tibshirani, 1993].

In section 2 and 3, the data and methodologies are discussed; results are covered in section 4; follow by a summary in section 5.

2. Data

A. Observational Data

The observational data are acquired from the National Climatic Data Center (NCDC) website based in Asheville, North Caroline. After filtering the data based on missing values, only 16 stations were useable for downscaling purposes. To ensure adequate comparison between different variables and locations, all data were standardized according to the mean and standard deviation. Since the untransformed variable does not following a Gaussian distribution, neither will the transformed data.

B. NCEP Reanalysis II Data

Potential predictors variables used for the training of the statistical model are obtained from the National Centers for Environmental Prediction (NCEP) Reanalysis II model runs for the time period 1979-2008. The model runs at a resolution of 2.5° latitude x 2.5° longitude on a global grid and nine variables at 17 different pressure levels are acquired. Due to a lack of large-scale observational data sets at various pressure levels (i.e., zonal and meridional winds), we must assume NCEP Reanalysis II data to be a gridded observation data set in order to allow for a statistical relationship to be built between large-scale NCEP and local-scale NCDC data.

C. GCM data

Hindcast (1979-2008) and future (2011-2040) daily precipitation data from 24 different GCMs and their various emission scenarios were obtained and analyzed from within the Environment Canada website. Complete GCM datasets were obtained through the World Data Center for Climate CERA Gateway.

3. Methods

A. Predictor Selection

Two different selection methods based on the correlation between the potential predictors and predictand have been employed and compared due to the fact the results from statistical downscaling are extremely sensitive to the choice of predictors. The first method is the Pearson correlation, which is the ratio of the covariance between potential predictors and predictand to the product of their standard deviations. The second and more robust is the Spearman rank correlation, which uses the rank of the raw data. The two correlation analyses were compared and the potential predictors which had the greatest correlation with the predictand (precipitation) from both methods are selected as inputs into the neural network.

B. MLP Topology and Learning Rule Selection

Artificial neural networks are extremely configurable with the commonly used configurations called topologies. In the current study, we applied the multilayer perceptron (MLP) topology based on its wide usage in current literature and relative ease to set up [Hsieh, 2009].

MLPs are trained by use of error-correction learning in which the network works to minimize the error between model output and desired output through repeated model runs and adjustment of model weights. The method by which the weights are adjusted is known as the learning rule and is a primary component of many ANN topologies.

Six different learning rules were tested by training the MLP network and keeping all components the same. The model output of extreme precipitation events was then compared to the observational extreme events by use of a 2x2 contingency table and various metrics calculated from the table. The learning rule which performed the best among the metrics was selected.

C. Division of Data for Model Use

The predictors and predictand data for each model run must be divided into three different groups. The training data consisted of 80% of the data. It is used in building the model and discovering the relationship between the predictors and predictand. To eliminate the model "memorizing" the training data, known as overfitting, 10% of the data is used as a cross-validation set. Overfitting results in the model not performing well with independent data outside the training phase. Lastly, 10% of the data must be kept separate from the training and cross-validation phase for use as an independent testing data set into the model after it has been built. In order to properly verify the testing phase, cross-testing was implemented which retrains the model ten times using a separate 10% of the data each time for each station [Hsieh, 2009]. This allows for the entire data set to be tested.

D. GCM Selection

Currently, there are ~24 different GCMs available for downscaling purposes. Each is made in a different region of the world and often is designed with that region's interests in mind. This has resulted in models that perform drastically different from one location to another. It is therefore important to determine which GCM of the 24 is appropriate for our downscaling variable and region.

A two-test approach suggested by the IPCC was applied to the suite of models in an attempt to determine the best model for precipitation on Oahu [IPCC, 2007]. The first is a baseline test which finds the absolute difference between the average daily observational precipitation during 1979-2008 and GCM precipitation hindcasts of the same time period. The GCMs are then ranked according to the resultant difference. The second test is considered a future projection (2011-2040) test which finds the absolute difference between the mean of all 24 models and each GCM. The GCMs are then ranked according to this resultant difference. The two ranked lists are compared to find an overall high ranked model and scenario.

4. Results

A. Predictors Selection

Overall, both the Pearson correlation and Spearman rank correlation analysis suggested four predictors out of all possible candidates. The predictors found best for use in downscaling extreme precipitation are: relative humidity at 850 hpa, the zonal wind component at 850 hpa, the meridional wind component at 1000 hpa, and sea-level pressure.

In many of the dry regions on Oahu, large synoptic events such as Kona storms have the largest influence on precipitation extremes. During such events, typical northeast trade winds change to moist southerly flow. The change in low level winds and moisture is reflected by the variables selected through the correlation analysis. [Timm and Diaz, 2009] showed near surface meridional wind to have the largest influence on island precipitation.

In regions where local-scale processes (i.e., orographic uplifting) dominate the rainfall patterns, correlation analysis did not suggest any adequate predictors for downscaling purposes. This is to be expected since it is the nature of statistical downscaling to find the relationship between local precipitation and large-scale processes. As a result, the 16 available stations were reduced to only seven. Of the seven, five are leeward dry stations and two are windward wet stations.

(a) Levenberg Marquardt learning rule		
	Obs (Yes)	Obs (No)
Forecast (Yes)	23	6
Forecast (No)	9	8722
(b) MLP extreme event performance		
	Obs (Yes)	Obs (No)
Forecast (Yes)	19	18
Forecast (No)	22	10891

Table 1: 1(a): The 2x2 contingency table for precipitation > 90th percentile during training of time period 1979-2008 for the Campbell Station. The Levenberg Marquardt learning rule is implemented in MLP. 1(b): The 2x2 contingency table for precipitation extremes of the MLP cross-testing data set during 1979-2008 at the Campbell Station.

B. Choice of Learning Rule and Performance Measures

Comparison of learning rules suggested, Levenberg-Marquardt, to be the only learning rule out of the six to have any skill at modeling extreme events. This is best represented in Table 1(a), which shows a 2x2 contingency table for the Levenberg-Marquardt learning rule during the training phase at the leeward Campbell station.

Several metrics are easily calculated from the contingency table in order to evaluate the model's forecast skill. Hit rate (HR) is used as an accuracy measure because it measure the proportion of observed events correctly forecasted. The worst possible value for HR is zero and the best is one. Frequency bias (FB) is used to evaluate the model's forecasting bias with an unbiased forecast of 1 and a value greater/less than one indicating the event was forecasted more/less often than observed. A commonly used skill score to evaluate forecast performance is known as the Peirce skill score (PSS) [WMO, 2002]. It was chosen because contributions made to the score by a correct no or yes forecast increases as the event is more or less likely, respectively. A perfect forecast would result in a PSS=1 while a random forecast results in

PSS=0. Another skill score used when the event occurs substantially less frequently than the nonoccurrence is the Gilbert Skill Score (GSS) with a perfect score of 1 and the worst score of 0. Finally, the Extreme Dependency Score has been calculated to assess the skill of the model at forecasting rare events. For a perfect forecast, EDS = 1 and for random forecasts EDS = 0. The 1(a) contingency table metrics have been summarized in the second column of Table 2.

Attribute	Learning Rule Training Performance	MLP Cross-Test Performance
HR	0.72	0.46
FB	0.91	0.90
PSS	0.72	0.46
GSS	0.60	0.32
EDS	0.88	0.75

Table 2: The scalar attributes of Hit Rate (HR), Gilbert Skill Score (GSS), Pierce Skill Score (PSS), Frequency Bias (FB), and Extreme Dependency Score (EDS) for contingency table 1a and 1b.

It is clear from the second column of Table 2 that the Levenberg-Marquardt learning rule outperforms the benchmark random forecasts.

C. MLP ANN Performance

Having selected the appropriate predictors, topology and learning rule for our study, it is possible to train and test the ANN performance. Table 1(b) demonstrates the model's skill through use of cross-testing at the Campbell Station. The metrics for Table 1(b) are given in the third column of Table 2. While it is clear that model skill decreases during the cross-testing as the sample size increases, the metrics still indicate an ability to outperform a random forecast. Model performance is further evaluated for the training and cross-testing phases through use of the Pearson correlation coefficient and RMSE skill score relative to a persistence model. These statistics are summarized in Table 3.

During the training stage four of five leeward stations exhibit significant correlation values at the 5% level between observations and model outputs. During the independent 'test' phase, three stations are statistically significant. The RMSE skill score in the last column of Table 3 suggests that the ANN is skillful at producing extreme precipitation events in leeward regions during the current climate.

Station	MLP	MLP	RMSE

	Training	Test	SS
Campbell	0.64*	0.29*	0.31
Honolulu	0.51*	0.21	0.28
Honolulu International Airport	0.72*	0.34*	0.34
Paiko	0.48*	0.27*	0.22
Punchbowl	0.29	0.17	0.08
Pali Golf	0.09	0.04	-0.71
Waimanalo Farm	0.05	0.03	-1.56

Table 3: MLP model performance for daily precipitation extremes in terms of correlation (second and third columns) and RMSE skill score (the last column) during 1979-2008. The top five stations are located in the dry region and the bottom two stations are located near the wet region of Oahu. * denotes the statistical significance of correlations at the 5% level.

D. GCM Selection and Simulation

The two-test approach outlined in Section 3D suggests the GCM ECHAM 5 scenario A2 as the most appropriate for downscaling daily precipitation for Oahu. Assuming the current day relationships between predictors and predictand hold valid into the future under climate change, the GCM predictor variables can be used as input into the MLP ANN after each run of the cross-testing procedure for each station. This creates a ten model ensemble at each station from which the final resultant statistics are calculated as the ensemble average. To assess the confidence interval of the given statistics, the BCa percentile resampling technique is employed [e.g., Efron and Tibshirani, 1993; Chu et al., 2009].

As an example, the Honolulu International Airport can illustrate the changes in precipitation extremes on leeward Oahu. During the current climate period (1979-2008), 47 heavy rainfall events are observed at this station while the model output shows 44 extreme events (Table 4). This model bias is found to occur at all leeward stations. During the future 30 years (2011-2040), the ANN predicts 55 extreme events. It is possible given the model bias, that the 55 events are slightly lower than reality. It is important to note that the 95% confidence intervals for the future climate shift to higher values relative to the current climate in both observations and model simulations. However, the spread of the intervals for the future climate are conservative relative to the current climate.

	Campbell	Honolulu International Airport	Honolulu (Observatory)	Paiko	Punchbowl
Extreme Frequency (1979-2008)	41	47	51	62	58
Confidence Interval (1979-2008)	[29,54]	[36,63]	[38,67]	[48,79]	[45,74]
Model Extreme Frequency (1979-2008)	37	44	46	55	50
Confidence Interval Model (1979-2008)	[24,51]	[32,59]	[33,62]	[43,72]	[38,65]
Model Extreme Frequency (2011-2040)	46	55	54	67	64
Confidence Interval Model (2011-2040)	[37,63]	[43,72]	[41,71]	[50,87]	[48,83]

Table 4: Extreme rainfall frequency for observed (1979-2008), current climate (1979-2008) from model, and future climate (2011-2040) from model. The corresponding 95% confidence interval of the storm frequency based on the BCa bootstrap resampling method is given in parenthesis.

Again, using Honolulu International Airport as an example and looking at rainfall intensity, the average value of 79 mm/day is observed. The model simulations indicate a dry bias towards lower values at a prediction of 73 mm/day. During the next 30 years, the average intensity of an extreme event is predicted to be 64 mm/day. This is lower than the average intensity of the actual observations and model simulation during the current climate. Combining results from both tables, our study suggest that in the next 30 years, the frequency of extreme events will increase but their mean intensity will decrease on leeward Oahu.

	Campbell	Honolulu International Airport	Honolulu (Observatory)	Paiko	Punchbowl
Average Extreme Intensity (mm/day) (1979-2008)	90.4	78.5	74.4	81.4	88.8
Confidence Interval (1979-2008)	[80.8,105.7]	[72.4,88.7]	[67.6,84.7]	[75.3,89.3]	[82.5,99.0]
Average Extreme Model Intensity (mm/day) (1979-2008)	84.2	73.4	65.9	74.7	79.6
Confidence Interval Model (1979-2008)	[74.3,97.8]	[66.1,85.07]	[58.0,77.6]	[68.9,83.2]	[71.4,90.9]
Average Extreme Model Intensity (mm/day) (2011-2040)	74.8	64.4	59.9	68.1	72.5
Confidence Interval Model (2011-2040)	[65.2,88]	[57.3,75.1]	[51.3,71.5]	[61.6,77.4]	[64.2,83.3]

Table 5: Same as Table 4 but for mean extreme rainfall intensity.

5. Conclusion

Heavy rainfall and flash floods are common in the Hawaiian Islands due to their steep terrain, orographic mechanisms, rain-producing weather systems, and abundant moisture supply. Given the socio-economic repercussions resulting from past storm events, it is of considerable interest to investigate changes in the frequency and intensity of heavy rainfall events in Hawaii, particularly for Oahu as it is the most populous island in Hawaii.

This study is based on observational station data on Oahu, NCEP/DOE reanalysis II data, and GCM data to project future changes in precipitation extremes via a artificial neural network using the MLP topology. Due to the limited availability of long-term and complete precipitation records, as well as local-scale processes which affect precipitation, stations selected in this study are restricted to only seven on Oahu. Using a large number of GCMs and their emission scenarios, the two-test approach recommended by IPCC reveals that the ECHAM5 A2 is the most appropriate in downscaling extreme precipitation events for Oahu. It is found that MLP networks performed better in drier areas. The MLP trained models are used together with ECHAM5 A2 data to provide estimates of the model's present-day climate and future climate.

There is a general agreement in key test statistics (e.g., the frequency of extreme events) between actual observations and GCM outputs under present-day conditions at all five leeward stations, although the model exhibits a small bias in underestimating both the frequency of storm occurrences and their mean intensity. For future projection (2011-2040), the model calls for higher number of extreme events but lower mean intensity relative to the present-day statistics. Considering the model bias, the rainstorm in the future would occur even more frequently than those indicated in Table 4 and its average intensity would be stronger than those given in Table 5. To provide a range of variability of the test statistics, a nonparametric BCa bootstrap technique is used for all three datasets (i.e., actual observations, GCM outputs from current climate, future GCM simulations).

References

Chu, P.-S., X. Zhao, Y. Ruan, and M. Grubbs, 2009: Extreme rainfall events in the Hawaiian Islands. *J. Appl. Meteor. Climatol*., 48, 502–516.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability*. Chapman & Hall, 456 pp.

Hsieh, W. W., 2009: *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*, Cambridge University Press, Cambridge, UK, 364 pp.

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc*., 79,1855–1870.

Intergovernmental Panel on Climate Change, 2007: Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Eds., S. Solomon et al., Cambridge Univ. Press, Cambridge, U. K., 1008 pp.

Kodama, K., and G.M. Barnes, 1997: Heavy rain events over the south-facing slopes of Hawaii: Antecedant conditions. *Wea. Forecasting*, 12, 347–367.

Norton, C., P.-S. Chu, and T.A. Schroeder, 2011: Projecting changes in future heavy rainfall events for Oahu, Hawaii: A statistical downscaling approach. J. Geophy. Res., in press.

Timm, O., H. F. Diaz, 2009: Synoptic-statistical approach to regional downscaling of IPCC 21st century climate projections: seasonal precipitation over the Hawaiian Islands. *J. Climate*, 22, 4261–4280.

WMO, 2002: Standardized verification system (SVS) for long-range forecasts (LRF). Manual on the GDPS. WMO No-485 Vol. 1, 5 pp.