Projecting changes in future heavy rainfall events for Oahu, Hawaii: A statistical downscaling approach

Chase W. Norton,[1] Pao-Shin Chu,[1] and Thomas A. Schroeder[1]

Received 14 January 2011; revised 3 June 2011; accepted 10 June 2011; published 14 September 2011.

[1] A statistical model based on nonlinear artificial neural networks is used to downscale daily extreme precipitation events in Oahu, Hawaii, from general circulation model (GCM) outputs and projected into the future. From a suite of GCMs and their emission scenarios, two tests recommended by the International Panel on Climate Change are conducted and the ECHAM5 A2 is selected as the most appropriate one for downscaling precipitation extremes for Oahu. The skill of the neural network model is highest in drier, leeward regions where orographic uplifting has less influence on daily extreme precipitation. The trained model is used with the ECHAM5 forced by emissions from the A2 scenario to simulate future daily precipitation on Oahu. A BCa bootstrap resampling method is used to provide 95% confidence intervals of the storm frequency and intensity for all three data sets (actual observations, downscaled GCM output from the present-day climate, and downscaled GCM output for future climate). Results suggest a tendency for increased frequency of heavy rainfall events but a decrease in rainfall intensity during the next 30 years (2011–2040) for the southern shoreline of Oahu.

Citation: Norton, C. W., P.-S. Chu, and T. A. Schroeder (2011), Projecting changes in future heavy rainfall events for Oahu, Hawaii: A statistical downscaling approach, *J. Geophys. Res.*, *116*, D17110, doi:10.1029/2011JD015641.

1. Introduction

[2] General circulation models (GCMs) play a pivotal role in the understanding of climate change. However, the limitations of GCMs restrict studies that result from their direct outputs due to coarse spatial resolution (~250 km). The need for site-specific climate change information requires a resolution finer than the capabilities of current GCMs. Statistical downscaling provides methods to obtain a finer resolution from the coarse GCM output.

[3] Statistical downscaling consists of finding an empirical relationship between large-scale atmospheric variables (predictors) and a small-scale variable (predictand or response variable). Finding the optimal relationship requires training of the model. Once the optimal relationship has been found the model must be validated to determine its ability to handle new and independent data. Careful consideration must be given when deciding how to choose the appropriate predictors, because statistical downscaling is extremely sensitive to the choice of predictors. Various downscaling methods have been used in literature, but the most commonly applied has been transfer functions. This method has four necessary conditions: (1a) there must be a strong relationship between large-scale predictors and the local-scale predictand; (2) the predictor must be well modeled; (3) the predictor parameters respond to a given perturbation in a similar manner as the predictand; and (4) the current relationship between the predictor and predictand must hold valid in the future under climate change [*Benestad et al.*, 2008].

[4] Principal component analysis, canonical correlation analysis, and linear regression techniques have been successfully employed to determine the statistical relationship between the predictors and the predictand [e.g., *Conway et al.*, 1996; *Crane and Hewitson*, 1998; *Schubert and Henderson-Sellers*, 1997]. These are linear methods and often the skill of these techniques is determined by the linearity of the variables studied. Recently, a study used linear techniques to analyze the relationship between ENSO, PNA and rainfall extremes over the Hawaiian Islands [*Elison Timm et al.*, 2011]. The results indicated small changes in rainfall extremes, but large uncertainties due to differences among climate models.

[5] Traditionally, linear statistical methods are used to explore the coupled variability between large-scale circulation features and local rainfall on monthly or seasonal time scales. This study focuses on extreme precipitation events based on daily records, which may not respond linearly to atmospheric forcings. For this reason, we applied a nonlinear method known as the neural network. This method offers many advantages, including the ability to find all possible connections between predictors, the ability to find the nonlinear relationship between predictors and response variable, ability to map input signal to desired response through supervised learning, multiple topologies to choose from based on application, and has been used extensively in academic research. However, there are disadvantages that must be understood and considered when analyzing the

[1]Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA.

0148-0227/11/2011JD015641

results of neural networks. By design they are "black boxes" and their output is difficult to interpret. This limits the research when trying to explain the cause of the results. Neural networks are prone to overfitting the data, which must be properly handled to provide reliable results. Neural networks are also computationally expensive and time consuming [*Tu*, 1996].

[6] *Hsieh and Tang* [1998] showed that artificial neural networks (ANNs) could provide an adequate method for handling nonlinear relationships commonly found in meteorology. *Cannon and Whitfield* [2002] had success using neural networks to find the complex nonlinear relationship between atmospheric variables and stream flows in British Columbia. The current study focuses on a nonlinear approach based on ANNs for statistically downscaling daily precipitation extremes on Oahu, Hawaii, for the time periods 1979–2008 (current climate) and 2011–2040 (future climate). Because the ordinary bootstrap procedure does not handle samples from heavy-tailed distributions very well, a BCa percentile bootstrap resampling method [*Efron and Tibshirani*, 1993] is used to assess the confidence interval of the test statistics (e.g., the frequency of heavy rainfall events).

[7] The objective of this study is to examine changes in the frequency and intensity of heavy precipitation events for Oahu, Hawaii. Section 2 describes Hawaii's climate and some recent heavy rainfall events as well as their damage. In section 3, observational, reanalysis and GCM data are discussed. Section 4 discusses the methodology for predictor selection. Section 5 addresses the results from the predictor selection, statistical model performance, and evaluates the change in precipitation extremes between 1979 and 2008 (i.e., current climate) and 2011–2040 (future climate).

2. Recent Heavy Rainfall Events and Hawaii's Climate

[8] The Hawaiian Islands are vulnerable to heavy rainfall and associated flooding from time to time [*Kodama and Barnes*, 1997; *Lyman et al.*, 2005]. For example, in late October 2004, Mānoa Valley in Oahu received 254 mm of rainfall within a 12 h period. Because Mānoa Stream swelled and topped its banks, the floodwater spilled with mud and debris onto the University of Hawai'i at Mānoa campus and its adjacent areas. This rain storm resulted in over 80 million USD worth of damage for the University of Hawai'i alone [*Chu et al.*, 2009]. From February to early April 2006, Hawaii was battered by rain. March 2006 was an exceptionally wet and record-breaking month in Hawaii. On 14 March 2006, the Ka Loko Dam on Kauai broke, releasing millions of gallons of water downstream in a gush. Seven people were killed and dozens of homes and properties were severely damaged. Multiple parties, including the owner of the dam, the State of Hawaii, and the County of Kauai faced liabilities for the disaster. In December 2008, an unusually intense storm developed in the vicinity of the Hawaiian Islands and produced several days of severe weather and the island of Oahu bore the brunt of the storm [*Smith*, 2009]. This rainstorm caused an estimated $50 million in damage and garnered a federal disaster declaration. In this study, Oahu is selected because it is the most populous island in the State of Hawaii and future water resource planning and management would have the largest impact in this region (Figure 1).

[9] The Hawaiian Islands have two distinct seasons, a dry (warm) season and a wet (cool) season. The dry season is from May to October and is dominated by trade winds. In the eastern North Pacific a quasi-stationary subtropical high advects northeasterly trade wind flows over the islands. This flow results in orographic lifting being the dominant mode of rainfall. During this time period, the location of maximum precipitation is on the windward slopes of the mountain ranges. Due to the rain shadow effect, the leeward side of the mountain ranges tends toward drier conditions. This region is characterized not only by low mean annual rainfall, ranging from 500 to 1000 mm, but also lower rainfall intensity associated with a specific return period [*Chu et al.*, 2009]. Besides trade winds, thermally driven diurnal circulations such as land-sea breezes also contribute to rainfall development by interacting with orographic uplifting and prevailing airflow [*Leopold*, 1949]. However, given the relatively small size of Oahu and its lower terrain height compared to the lifting condensation level, orographic lifting alone is inadequate to produce rainfall [*Hartley and Chen*, 2010]. Factors important for abundant orographic rainfall are the existence of cumulus clouds over the adjacent ocean, a deep moist layer, and sufficient moisture content upstream of the island under strong trade wind conditions. Extreme precipitation events are commonly a result of tropical cyclone activity during the warm season.

[10] The wet season extends from November to March or April. At this time, normal trade wind patterns are occasionally interrupted by midlatitude fronts, Kona storms, and upper level disturbances [*Schroeder*, 1993; *Chu et al.*, 1993]. Kona storms are unique to the Hawaiian Islands. They result from cutoff lows in the upper level westerlies and commonly cause widespread extreme precipitation events that last days or even weeks. Due to the synoptic nature of these events and their preferred geographical locations, extreme precipitation is not limited to the ordinary east facing windward slopes.

3. Data and Model Description

[11] The observational, reanalysis, and GCM data are obtained for the specified time period. The observational and National Centers for Environmental Prediction (NCEP) data are used to train the statistical models. It is assumed that NCEP data represent observational data well enough to be considered a gridded observational data set. This assumption allows for a statistical relationship to be formed between observed local-scale precipitation and large-scale atmospheric reanalysis variables for the same given time period. Single-site downscaling with ANNs, as opposed to multisite methods, is found to be more applicable for daily precipitation due to the complex terrain on Oahu (Figure 1). The commonly applied multilayer perceptron (MLP) topology is employed based on its wide usage in current literature [*Hsieh*, 2009; *Mendes and Marengo*, 2010].

[12] Oahu is of volcanic origin and comprised of two parallel mountain ranges, Koolau and Waianae, separated by a broad interior valley known as Central Oahu (Figure 1). The mountain ranges are the eroded remains of ancient shield volcanoes formed millions of years ago. The Koolau

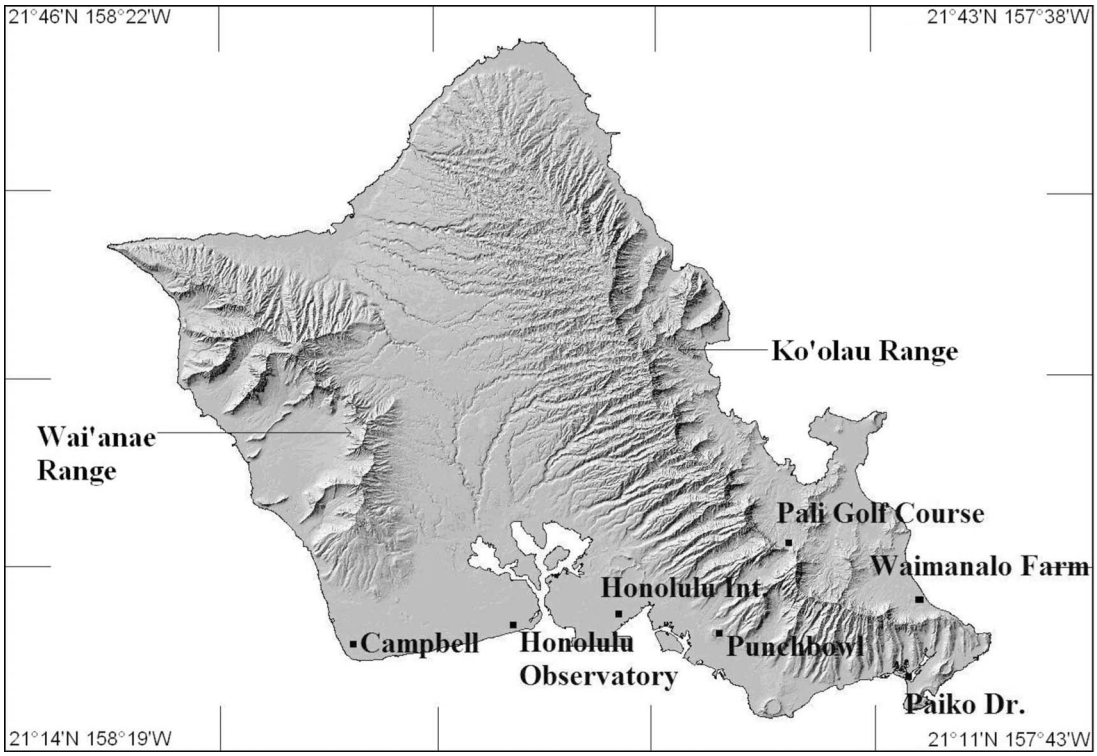

Figure 1. Map of stations on Oahu used for this study.

Mountain range spans the length of the eastern coast, roughly 45 km long with peaks varying from 500 to 960 m high. It is aligned almost 90 degrees in the direction of northeast trade winds, which results in wetter conditions and faster erosion. The Waianae Mountain range spans from 470 to 1,200 m and is on the western side of the island. This range separates central Oahu from the Leeward coast. Its location relative to the Koolau range and normal trade wind flow means it is less eroded and often drier.

3.1. Data

3.1.1. Observational Data

[13] Observational daily precipitation data are obtained through the National Climatic Data Center (NCDC), in Asheville, North Carolina. When obtaining the data, only stations with at least 30 years of daily precipitation data are used. Further, a three and five rule is applied. This rule states that if three consecutive days are missing then the month must be removed or five nonconsecutive days are missing in a month then the month must be removed. Oahu has a total of 103 precipitation gauges, and only 16 are usable after the filter process. All data are then standardized according to the mean and standard deviation [*Hewitson and Crane*, 2006]. This transformation ensures that the resulting precipitation series will be dimensionless and exhibits a mean of zero and a standard deviation of one so the variation on wet windward stations is comparable to the dry leeward stations. It also allows for a direct comparison of atmospheric variables with different units to the observed standardized precipitation. It should be noted that the transformed data will not follow a Gaussian distribution since the untransformed variable does not either. *Hewitson and Crane* [2006] used the standardized anomalies for statistical downscaling over South Africa.

3.1.2. NCEP Reanalysis II Data

[14] NCEP reanalysis II daily data are obtained for the gridded predictor data set and used to train the model. The data set consists of nine candidates of atmospheric variables at 17 different pressure levels. The model resolution is 2.5 degree latitude × 2.5 degree longitude on a global grid. Reanalysis II data have shown better representation of winter time precipitation and tropical precipitation when compared with reanalysis I data [*Kanamitsu et al.*, 2002]. Due to the location of the Hawaiian Islands in the tropics and because most of the extreme precipitation events occur in the cooler season, reanalysis II data are used. Both NCEP and GCM data are standardized according to the same methods for observed data.

3.1.3. GCM Data

[15] We analyzed 24 GCMs and their emission scenarios to determine which would be appropriate in downscaling daily precipitation over Oahu. The two approaches suggested by the International Panel on Climate Change (IPCC) are applied to each model [*Pachauri and Reisinger*, 2007]. The first approach is a baseline test that compares the average observational precipitation during 1979–2008 with GCM back projections of the same time period. An island average for daily precipitation is calculated to compare with a grid located closest to Oahu from each model. The 24 models and their emission scenarios are ranked based on the absolute difference between island observed average precipitation and the precipitation from each GCM. This approach can be regarded as weighting the model's bias under current climate conditions. It is tacitly assumed that

the GCM that most accurately represents local precipitation is also the most adequate model that can simulate precipitation related large-scale circulation features (i.e., predictors) well. The second approach is a future projection (2011–2040) test that compares the grid box located closest to Oahu from each GCM model in order to filter outliers. GCMs that lay outside of ±1 standard deviation of the 24 model mean are discarded. Those that lay inside ±1 standard deviation are regarded as close to the consensus (i.e., convergence) and are thus ranked according to absolute difference from the mean.

[16] The aforementioned two tests are then compared to find an overall high ranked model among all 24 GCMs and their scenarios, and are similar to the reliability ensemble average method [*Giorgi and Mearns*, 2002]. For our study, the combined bias and convergence approach for multimodel evaluation suggests that GCM model ECHAM5 and scenario A2 have one of the smallest differences for the current climate (1979–2008) and also for the future climate (2011–2040). Multiple grid boxes located around Oahu are also tested and have similar results as the grid box closest to Oahu. In the following, we will introduce the basics of the neural networks used in this study.

3.2. MLP Topology

[17] Unlike other topologies, MLP networks have no restriction on hidden layers. They are considered to be a universal function approximator. *Hornik et al.* [1989] showed that one hidden layer could approximate any continuous function to a degree of accuracy, so for this reason only one hidden layer is used (Figure 2). Within the hidden layer the optimal number of processing elements (PEs) is found to be 25. If there are too many PEs the network begins to 'memorize' the training data and is unable to model new data outside of the training phase. On the other hand, if there are too few PEs the network is not able to recognize the patterns necessary for an accurate model [*Principe et al.*, 1999]. Therefore, a compromise of 25 PEs is chosen.

[18] MLP networks are trained through error-correction learning with the most common being the generalized delta rule or backpropagation, which uses the methodology of gradient-descent learning [*McClelland and Rumelhart*, 1986]. The goal when using backpropagation is to minimize the distance between the desired response and model output. This requires the desired response be known prior to model training in order to allow for comparison and model improvement. The model initiates by assigning weights to the input vectors (predictors) and applying nonlinear stepwise transformations. The output is compared to the desired response (predictand) and an error between them is calculated. The errors are back signaled through the topology and the weights are adjusted according to the error. This process is repeated until a termination criterion is met. The methodology for how to adjust the weights on the input vectors is known as a learning rule and is a key component to ANNs. Six different learning rules are tested in this study: step; momentum; delta-bar-delta; quickprop; conjugate gradient; and Levenberg-Marquardt.

[19] For each model run the data are divided into three separate groups. The training data (80% of the data) are used to build the model and discover the relationship between predictand and predictors. It is required that the training data set be sufficiently large enough to discover the signal. The cross-validation set (10% of the data) is used to eliminate overfitting, which is a primary concern in neural networks. If the network overfits the data, then it runs the risk of not being able to handle independent data well outside of the training phase. In our study, we attempt to prevent this through the use of the cross-validation data set.

[20] Cross validation is a generalization of the common procedure of omitting a few observations from the data, reconstructing the model using the remaining data, and then making predictions for the omitted cases. At the end of each training run, the cross-validation data are inputted into the model. The root mean squared error (RMSE) is then calculated for both training and cross-validation sets. If the RMSE of the training set decreases, but the cross-validation RMSE begins to increase then the model is considered overfitting the training data. Essentially, when this happens the model is 'memorizing' the training data, which degrades the ability to model new data. When this occurs the model run is terminated.

[21] The testing set (10%) is used to determine how well the model responds to new data that are separate from the training and cross-validation phase. To ensure proper verification of the models, cross testing was performed to allow extending the testing of the entire data set [*Hsieh*, 2009]. For cross testing, the trained model is tested using the 10% testing data set. The model is then retrained and tested using a separate 10% of the data. This process is repeated successively by changing the portion of the data that has been included from the model testing. By doing so, one can test the entire data set. [e.g., *Chu et al.*, 2010a; *Hsieh*, 2009]. After deliberately dividing the entire data set into training, cross-validation, and testing sets the next task is to perform predictor selection, because there are always more potential predictors available than can be used and some predictors inherently carry redundant information.

4. Methodology of Predictor Selection

[22] Two methods are employed to determine the best predictor combination for downscaling daily precipitation: Pearson correlation analysis and Spearman rank correlation. A Pearson correlation is the ratio of the covariance between the input and desired data to the product of their standard deviations and is most commonly used to reveal the linear association between two variables. Spearman rank correlation is a robust and resistant alternative to the Pearson correlation. It is simply the Pearson correlation coefficient but computed using the rank of the raw data instead of the actual data. The sign of the ranked data is an indicator of the direction of the relationship between the predictand and the predictor and the magnitude of the rank coefficient indicates the importance of the relationship.

5. Results

[23] Results have been separated into four subsections. Section 5.1 addresses the predictor selection. The choice of learning rule is described in section 5.2. Section 5.3 describes the model performance during training and testing phases. Section 5.4 addresses the results from inputting GCM data during the time period 2011–2040 into the trained neural network model.

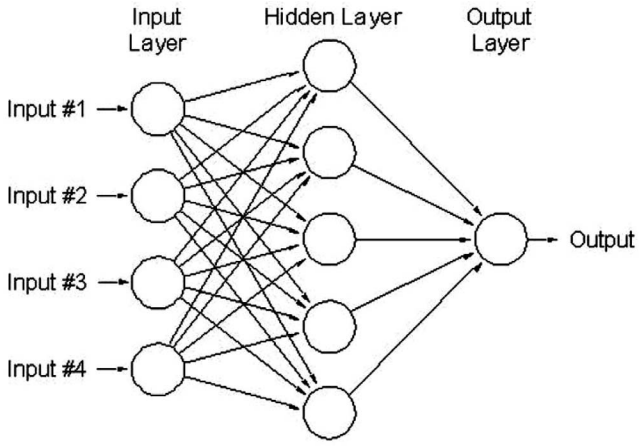

Figure 2. A representation of an artificial neural network with an input layer, a hidden layer, and an output layer.

5.1. Predictor Selection

[24] Both the Pearson correlation and Spearman rank correlation analysis between observations and NCEP data during 1979–2008 suggest four predictors out of all the possible candidate predictors. The predictors are: relative humidity at 850 hPa, zonal wind component at 850 hPa, meridional wind component at 1000 hPa, and sea level pressure.

[25] For Oahu, the two mountain ranges have a significant influence on daily precipitation. In the dry leeward regions where trade wind orographic lifting is not the dominant mode of extreme precipitation, correlation values from Pearson and Spearman rank are generally higher than those from windward mountain ranges. In these dry regions large synoptic events such as Kona storms have a stronger influence on precipitation extremes. The Kona storm is typically a nonfrontal low pressure system that comes in close proximity to the Hawaiian Islands. It brings moist southerly or southwesterly flows with occasional heavy precipitation to the normally dry regions. This change in the circulation from the common northeast trade wind days is reflected by the changes in low level winds and moisture possibly captured by the aforementioned four predictors. On the seasonal time scale, *Timm and Diaz* [2009] showed that the near surface meridional winds have the largest effect on island precipitation. For the leeward stations the same four predictors will be used for subsequent downscaling.

[26] The purpose of statistical downscaling is to find the empirical relationship between large-scale atmospheric variables and local precipitation. This relationship, however, is a poor fit for Oahu's windward stations where local-scale processes (i.e., orographic uplifting) strongly affect precipitation. Consequently, the 16 available stations described in section 3.1.1 are reduced to only seven for downscaling (Figure 1). Table 1 lists the basic statistics of daily precipitation from these seven stations during 1979–2008. The first five are on the leeward side of the Oahu and the next two are on the windward side. As expected, the mean and standard deviation of daily precipitation are lower at leeward gauges relative to windward stations. The frequency of wet days, defined as the number of days with more than 1 mm precipitation, is also generally lower on the leeward stations. However, when it comes to wet days' mean precipitation, there is no substantial difference among all seven stations. For example, the mean precipitation of wet days at Campbell is comparable to those at the two windward sites. This suggests that for the leeward side, although the frequency of wet days is a lot lower, the amount of mean precipitation received during wet days is roughly the same as the windward area. The standard deviation of precipitation during wet days is generally lower at leeward regions.

5.2. Choice of Learning Rule and Performance Measures

[27] For this study the learning rule, Levenberg-Marquardt, is found to be best at modeling extreme events. It is a higher-order adaptive algorithm known for minimizing the mean squared error (MSE) of a neural network. In terms of learning, it attempts to approximate the matrix of the second derivatives of the performance surface. This determines the best direction to adjust the weights in order to achieve a lower error. Extreme events are here defined as those that exceed the 90th percentile of daily rainfall distribution [*Chu et al.*, 2009].

[28] Table 2a shows a case using this learning rule during the training phase for the Campbell station. In this leeward region, heavy rainfall events are uncommon and often occur due to large-scale circulation conditions. Forecast results for ANN in categories (i.e., yes or no) are evaluated. Because heavy rainfall events are rare, Table 2a is dominated by a correct 'no' forecast.

[29] Verification for categorical forecasts is easier to understand and several measures are readily available to evaluate forecast skill. The hit rate (HR) is used as an accuracy measure. Hit rate measures the proportion of observed events correctly forecasted. The worst possible value for the HR is zero and the best is one. The frequency bias (FB) is calculated to evaluate the ratio of yes forecasts to the number of yes observations. A value of one would indicate an unbiased forecast, while a value greater/less than one would indicate the event was forecasted by the models more/less often than the observed. Moreover, one of the most frequently used skill scores for summarizing square

Table 1. Mean and SD of the Seven Stations[a]

	Campbell	Honolulu International Airport	Honolulu (Observatory)	Paiko	Punchbowl	Pali Golf	Waimanalo Farm
Mean (mm/d)	1.25	1.20	1.23	1.91	2.37	5.13	3.03
SD	7.15	6.49	6.60	7.71	7.11	15.10	11.35
Frequency of wet days	1422	1497	1530	2694	3742	5742	3539
Wet day mean (mm/d)	9.47	8.65	8.72	7.67	6.89	9.82	9.33
SD wet days	17.90	15.85	15.94	14.26	10.98	19.98	18.70

[a]Key statistics (mean and standard deviation) for the stations used in this study. The first five stations are located on the leeward and the last two stations are located on the windward Oahu. Unit for mean and standard deviation (SD) is mm/d. Unit for the frequency of wet days is days.

Table 2a. The 2 × 2 Contingency Table for Precipitation > 90th Percentile During Training of Time Period 1979–2008 for the Campbell Station[a]

	Observed (Yes)	Observed (No)
Forecast (Yes)	23	6
Forecast (No)	9	8722

[a]The Levenberg Marquardt learning rule is implemented in MLP.

contingency table is known as the Peirce skill score (PSS), which is chosen over other skill measures because the contributions made to the score by a correct no or yes forecast increases as the event is more or less likely, respectively. PSS is similar to the Heidke skill score (HSS) except that the imagined random reference forecasts in the denominator are constrained to be unbiased. For perfect forecasts, PSS = 1 and for random forecasts the score is zero. Note that PSS and HSS are standard metrics to evaluate forecast performance [*World Meteorological Organization*, 2002]. Because the event to be forecast occurs substantially less frequently than the nonoccurrence (Table 2a), the Gilbert Skill Score (GSS) is also used. It is often referred to as the ratio of success [*Wilks*, 2006]. The best value for GSS is one and the worst is zero.

[30] More recently, the Extreme Dependency Score (EDS) has been employed as an assessment of the skill the models have with forecasting rare events [*Stephenson et al.*, 2008; *Ferro*, 2007; C. A. T. Ferro and D. B. Stephenson, Extremal Dependence Indices: Improved verification measures for deterministic forecasts of rare binary events: Revised, http://empslocal.ex.ac.uk/people/staff/ferro/Publications/edi.pdf]. It is calculated from the equation: EDS = (2*log((a + c)/n)/(log(a/n))) − 1, where a is number of correct forecasts, c is the number of times the event occurred but was not forecast, and n is the total number of forecasts. For perfect forecasts, EDS = 1 and for random forecasts EDS = 0. The various skill scores from Table 2a are shown in the "Learning Rule Training Performance" column of Table 3. Clearly, the ANN model presented in this study outperforms the benchmark random forecast. It is found that the other five learning rules significantly underestimate the extremes.

5.3. MLP ANN Performance

[31] To further demonstrate the model's skill, Table 2b displays the 2 × 2 contingency table for the cross-testing data set at the Campbell station. This site is chosen as an example because the MLP topology is found to work rather well in both training and independent test stages (Table 4). For this entire data set, the correct "no" forecast category again dominates in Table 2b. The metrics for Table 2b are shown in the "MLP Cross-Test Performance" column of Table 3. While the model's skill decreases from the training period to cross testing as the sample size increases, the metrics still show the model's ability to outperform the random forecast. The MLP model performance for daily precipitation extremes during 1979–2008 is further evaluated in terms of the Pearson correlation coefficient and RMSE skill score relative to a persistence model (Table 4). The skill score based on the RMSE was calculated using the formula: $SS_{persistence} = 1 - (RMSE_{neural\ network}/RMSE_{persistence})$.

Table 2b. The 2 × 2 Contingency Table for Precipitation Extremes of the MLP Cross-Testing Data Set During 1979–2008 at the Campbell Station

	Observed (Yes)	Observed (No)
Forecast (Yes)	19	18
Forecast (No)	22	10,891

Table 3. Scalar Attributes for the Contingency Tables[a]

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

[a]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 Tables 2a and 2b.

[32] During the training stage, four of five leeward stations exhibit significant correlation values at the 5% level between observations and model outputs (Table 4). The Punchbowl station is marginally significant. For two windward stations (Pali Golf and Waimanalo Farm), the results from the model exhibit no linear association between observations and model simulations.

[33] During the independent 'test' phase, correlations are generally found to be high for stations that had high correlations during the training stage (Table 4). In this case, three of the five MLP models are statistically significant at drier stations. As shown in the "RMSE SS" column of Table 4, the skill score based on RMSE supports the notion that ANN is skillful in producing extreme precipitation events given the selected predictors (i.e., low-level moisture content, low-level winds, and sea level pressure) during the current climate condition. These values suggest that the MLP network is able to model extreme events reasonably well at the drier stations. Wetter station data tend to be noisy and MLP networks model toward the mean of precipitation rather than finding the patterns of extremes. Having tested the performance of ANNs using data for the current climate (1979–2008), the next task is to estimate changes in the frequency and intensity of precipitation extremes for the future thirty years starting from 2011. For this task, the projection of future climate from a GCM is used.

Table 4. MLP Correlation Performance[a]

Station	MLP Training	MLP Test	RMSE SS
Campbell	0.64[b]	0.29[b]	0.31
Honolulu	0.51[b]	0.21	0.28
Honolulu International Airport	0.72[b]	0.34[b]	0.34
Paiko	0.48[b]	0.27[b]	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

[a]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.
[b]Statistical significance of correlations at the 5% level.

Table 5. Frequency of Extreme Events[a]

	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]

[a]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 brackets.

5.4. Future Projection From GCM Simulations

[34] The two-test approaches outlined in section 3.1.3 suggest that for Oahu daily precipitation the GCM ECHAM5 A2 could be used. Assuming that the present-day relationships discovered will hold into the future, the GCM predictor variables are used as inputs into the statistical model built for the MLP stations after each run of the cross-testing procedure. This allows for a ten model ensemble to be made and the result is determined from the ensemble average of these model runs. In order to assess the confidence interval of the test statistics, the frequency and intensity of precipitation extremes from both the actual observations and model outputs are calculated using the BCa percentile resampling technique [e.g., *Efron and Tibshirani*, 1993; *Gilleland*, 2010; *Jolliffe*, 2007; *Chu et al.*, 2009]. The BCa method is an extension of the commonly used percentile method which simply uses $\alpha/2$ and $1 - \alpha/2$ percentiles of the bootstrap distribution to define the interval [e.g., *Chu*, 2002]. Here α is the common 5% level of significance. The conventional method works well when the resulting statistics are unbiased and have a symmetric sampling distribution. If statistics are biased, the common percentile method would amplify the bias. To overcome this problem, the BCa method corrects the common percentile interval for bias and skewness. In this study, 10,000 bootstrap replications are generated using the BCa percentile method.

[35] As an example, the Honolulu International Airport is used to illustrate changes in precipitation extremes on leeward Oahu. During the 1979–2008 period, 47 heavy rainfall events are observed at this station and there is a 95% confidence that their true values lie in the interval between 36 and 63 (Table 5). For the same period and the identical threshold for defining extreme events, the model output shows 44 extreme events. Thus, the ECHAM5 model slightly underestimates the frequency of events under current climate conditions. This model bias is also found at four other dry stations. For the next 30 years (2011–2040), the model output predicts 55 extreme events at the Honolulu International Airport. Note that the 95% confidence intervals for the extreme events in the future climate shift to higher values than those in the current climate from both actual observations and model simulations. However, the spread of the 95% confidence intervals for future climate is conservative relative to the current climate. Given the likelihood of GCM bias in simulating the number of precipitation extremes in the present-day climate (47 versus 44), it is possible that the 55 events at the Honolulu International Airport projected in the future also are slightly biased toward lower frequency than reality. In principle, this also applies to the other four sites.

[36] For rainfall intensity (Table 6), the average value is 79 mm/d at the Honolulu International Airport from actual observations, with 95% confidence intervals between 72 and 89 mm/day. This is in contrast to 73 mm/d from model simulations during 1979–2008. Compared to observations, the model exhibits a slightly dry bias, which is also prevalent at the other four stations. On average, the model simulates 7 mm less rainfall per day for all five stations. During 2011–2040, the average intensity of an extreme event at the Honolulu International Airport is predicted to be 64 mm/d. This projected average precipitation intensity is lower than those from the actual observations and model simulations under the current climate conditions. This decrease is also found at the other four stations. Among the five leeward stations, the decrease in average precipitation intensity from the current observation to the future is largest at Punchbowl, from the observed 89 mm/d in the current climate to 73 mm/d in the future. Combining results from changes in the frequency and intensity of precipitation extremes (Tables 5 and 6), our study suggests that in the next 30 years, the frequency of extreme events will increase but their mean intensity will decrease on leeward Oahu.

[37] In analyzing observed rainfall records from 1950s to 2007, *Chu et al.* [2010b] recently noted a prevailing downward trend in daily rainfall intensity on Oahu. In

Table 6. Intensity of Extreme Events[a]

	Campbell	Honolulu International Airport	Honolulu (Observatory)	Paiko	Punchbowl
Average extreme intensity 1979–2008 (mm/d)	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 1979–2008 (mm/d)	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 2011–2040 (mm/d)	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]

[a]Same as Table 5 but for mean extreme rainfall intensity.

particular, the decreasing trend at the Honolulu International Airport since 1950s is statistically significant at the 5% level using a nonparametric Mann-Kendall test and Sen's method. Therefore, the projected decrease in mean storm rainfall intensity in Table 6 is consistent with what was observed during the past 60 years. If the model bias is taken into account, the increase in the frequency of extreme events in the future would perhaps be even higher than the model's projection (e.g., 55 events at the Honolulu International Airport) and the average intensity will likely be stronger than the projected (e.g., 64 mm/d at the Honolulu International Airport) for the southern shoreline of Oahu.

6. Conclusion

[38] 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. They have caused multimillion dollars damage to homes, properties, roads, agriculture, and other sectors. Environmentally, heavy rainfall and runoff events in Hawaii, which are likely to cause slope and coastal erosion, pollutant discharges to nearshores, coral reef degradation, among others, are expected to change as the planet Earth has been undergoing an unprecedented warming process since the Industrial Revolution. Given the socioeconomic 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.

[39] 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 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. The application of ANNs results in relatively high correlation values (e.g., Table 4) for daily precipitation in both the training and independent test stages. 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.

[40] 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 5 and its average intensity would be stronger than those given in Table 6. To provide a range of variability of the test statistics, a nonparametric BCa bootstrap technique is used for all three data sets (i.e., actual observations, GCM outputs from current climate, future GCM simulations).

[41] The results presented in this study may benefit many agencies who are concerned with floods and relevant policy making in the face of climate change. For instance, changes in rainstorm intensity may be a serious concern for the Pearl Harbor aquifer as precipitation is the primary water resource for streams and groundwater supply. Because many people live near this aquifer and most of their water use is drawn from the freshwater lens buried underground, a projected decrease in precipitation extremes in the future would inevitably alter local hydrology and management practices for this aquifer. Due to the effect of the mountains on Oahu, areas with high average daily precipitation did not perform as well as leeward regions. Further research in dynamical downscaling would provide more insight into the physical processes causing the change in extreme events with increased greenhouse gases.

[42] Many tropical Pacific islands are experiencing rapid population growth that places an increasing demand on water for drinking, food production, recreation, and other needs. Further compounding this problem is the great variability of interannual and interdecadal precipitation in this region [e.g., *Chu and Chen*, 2005]. It is hoped that the method demonstrated in Hawaii would be of value to other tropical Pacific islands in projecting future precipitation extremes, which would be vital to various governments for future water resources planning and management.

[43] **Acknowledgments.** Thanks are owed to May Izumi for her excellent editing services and to John Marra, NOAA Pacific Regional Climate Services Director, as well as for the comments from the two anonymous reviewers which improved the quality and presentation of this paper. This report is prepared by the Joint Institute for Marine and Atmospheric Research, University of Hawai'i, under award NA090AR4320075 from the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce. This study is also partially funded by a grant/cooperative agreement from the NOAA, Project M/PD-1, which is sponsored by the University of Hawai'i Sea Grant College Program, SOEST, under institutional grant NA09OAR4170060 from NOAA Office of Sea Grant, Department of Commerce. UNIHI-SEAGRANT-JC-11-01. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies.

References

Benestad, R., D. Chen, and I. Hanssen-Bauer (2008), *Empirical Statistical Downscaling*, 228 pp., World Sci., Singapore.

Cannon, A. J., and P. H. Whitfield (2002), Downscaling recent streamflow condition in British Columbia, Canada using ensemble neural network models, *J. Hydrol.*, *259*, 136–151, doi:10.1016/S0022-1694(01)00581-9.

Chu, P.-S. (2002), Large-scale circulation features associated with decadal variations of tropical cyclone activity over the central North Pacific, *J. Clim.*, *15*, 2678–2689, doi:10.1175/1520-0442(2002)015<2678:LSCFAW>2.0.CO;2.

Chu, P.-S., and H. Chen (2005), Interannual and interdecadal rainfall variations in the Hawaiian Islands, *J. Clim.*, *18*, 4796–4813, doi:10.1175/JCLI3578.1.

Chu, P.-S., A. J. Nash, and F. Porter (1993), Diagnostic studies of two contrasting rainfall episodes in Hawai'i: Dry 1981 and wet 1982, *J. Clim.*, *6*, 1457–1462, doi:10.1175/1520-0442(1993)006<1457:DSOTCR>2.0.CO;2.

Chu, P.-S., X. Zhao, Y. Ruan, and M. Grubbs (2009), Extreme rainfall events in the Hawaiian Islands, *J. Appl. Meteorol. Climatol.*, *48*, 502–516, doi:10.1175/2008JAMC1829.1.

Chu, P.-S., X. Zhao, C.-H. Ho, H.-S. Kim, M.-M. Lu, and J.-H. Kim (2010a), Bayesian forecasting of seasonal typhoon activity: A track-pattern-oriented categorization approach, *J. Clim.*, *23*, 6654–6668, doi:10.1175/2010JCLI3710.1.

Chu, P.-S., Y. R. Chen, and T. A. Schroeder (2010b), Changes in precipitation extremes in the Hawaiian Islands in a warming climate, *J. Clim.*, *23*, 4881–4900, doi:10.1175/2010JCLI3484.1.

Conway, D., R. L. Wilby, and P. D. Jones (1996), Precipitation and air flow indices over the British Isles, *Clim. Res.*, *7*, 169–183, doi:10.3354/cr007169.

Crane, R. G., and B. C. Hewitson (1998), Doubled CO_2 precipitation changes for the Susquehanna basin: Down-scaling from the genesis general circulation model, *Int. J. Climatol.*, *18*, 65–76, doi:10.1002/(SICI)1097-0088(199801)18:1<65::AID-JOC222>3.0.CO;2-9.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, 456 pp., Chapman and Hall, London.

Elison Timm, O., H. F. Diaz, T. W. Giambelluca, and M. Takahashi (2011), Projection of changes in the frequency of heavy rain events over Hawaii based on leading Pacific climate modes, *J. Geophys. Res.*, *116*, D04109, doi:10.1029/2010JD014923.

Ferro, C. A. T. (2007), A probability model for verifying deterministic forecasts of extreme events, *Weather Forecast.*, *22*, 1089–1100, doi:10.1175/WAF1036.1.

Gilleland, E. (2010), Confidence intervals for forecast verification, *NCAR Tech. Note NCAR/TN-479+STR*, 71 pp., Natl. Cent. for Atmos. Res., Boulder, Colo. (Available at http://nldr.library.ucar.edu/collections/technotes/asset-000-000-000-846.pdf.)

Giorgi, F., and L. O. Mearns (2002), Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method, *J. Clim.*, *15*, 1141–1158, doi:10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2.

Hartley, T., and Y.-L. Chen (2010), Characteristics of summer trade-wind rainfall over Oahu, *Weather Forecast.*, *25*, 1797–1815, doi:10.1175/2010WAF2222328.1.

Hewitson, B. C., and R. G. Crane (2006), Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa, *Int. J. Climatol.*, *26*, 1315–1337, doi:10.1002/joc.1314.

Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer feedforward networks are universal approximators, *Neural Networks*, *2*, 359–366, doi:10.1016/0893-6080(89)90020-8.

Hsieh, W. W. (2009), *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*, 364 pp., Cambridge Univ. Press, Cambridge, U. K., doi:10.1017/CBO9780511627217.

Hsieh, W. W., and B. Tang (1998), Applying neural network models to prediction and data analysis in meteorology and oceanography, *Bull. Am. Meteorol. Soc.*, *79*, 1855–1870, doi:10.1175/1520-0477(1998)079<1855:ANNMTP>2.0.CO;2.

Jolliffe, I. T. (2007), Uncertainty and inference for verification measures, *Weather Forecast.*, *22*, 637–650, doi:10.1175/WAF989.1.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP-DOE AMIP-II Reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, *83*, 1631–1643, doi:10.1175/BAMS-83-11-1631.

Kodama, K., and G. M. Barnes (1997), Heavy rain events over the south-facing slopes of Hawaii: Attendant conditions, *Weather Forecast.*, *12*, 347–367, doi:10.1175/1520-0434(1997)012<0347:HREOTS>2.0.CO;2.

Leopold, L. B. (1949), The interaction of trade wind and sea breeze, Hawaii, *J. Meteorol.*, *6*, 312–320, doi:10.1175/1520-0469(1949)006<0312:TIOTWA>2.0.CO;2.

Lyman, R. E., T. A. Schroeder, and G. M. Barnes (2005), The heavy rain event of 29 October 2000 in Hana, Maui, *Weather Forecast.*, *20*, 397–414, doi:10.1175/WAF868.1.

McClelland, J. L., and D. E. Rumelhart (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. II, 1190 pp., MIT Press, Cambridge, Mass.

Mendes, D., and J. A. Marengo (2010), Temporal downscaling: A comparison between artificial neural network and autocorrelation techniques over the Amazon Basin in present and future climate change scenarios, *Theor. Appl. Climatol.*, *100*, 413–421, doi:10.1007/s00704-009-0193-y.

Pachauri, R. K., and A. Reisinger (Eds.) (2007), IPCC Climate Change 2007: Synthesis report, report, 104 pp., Intergov. Panel on Clim. Change, Geneva, Switzerland.

Principe, J. C., N. R. Euliano, and W. C. Lefebvre (1999), *Neural and Adaptive Systems: Fundamentals Through Simulations*, 1st ed., 672 pp., John Wiley, Hoboken, N. J.

Schroeder, T. A. (1993), Climate controls, in *Prevailing Trade Winds*, edited by M. Sanderson, pp. 12–36, Univ. of Hawai'i Press, Honolulu.

Schubert, S., and A. Henderson-Sellers (1997), A statistical model to downscale local daily temperature extremes from synoptic-scale atmospheric circulation patterns in the Australian region, *Clim. Dyn.*, *13*, 223–234, doi:10.1007/s003820050162.

Smith, N. P. (2009), Dynamic and thermodynamic contributions to the evolution of a rapidly intensifying Kona low, M.S. thesis, 99 pp., Dep. of Meteorol., Univ. of Hawai'i at Mānoa, Honolulu.

Stephenson, D. B., B. Casati, C. A. T. Ferro, and C. A. Wilson (2008), The extreme dependency score: A non-vanishing verification score for deterministic forecasts of rare events, *Meteorol. Appl.*, *15*, 41–50, doi:10.1002/met.53.

Timm, O., and H. F. Diaz (2009), Synoptic-statistical approach to regional downscaling of IPCC 21st century climate projections: Seasonal precipitation over the Hawaiian Islands, *J. Clim.*, *22*, 4261–4280, doi:10.1175/2009JCLI2833.1.

Tu, J. V. (1996), Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, *J. Clin. Epidemiol.*, *49*, 1225–1231, doi:10.1016/S0895-4356(96)00002-9.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 627 pp., Academic, San Diego, Calif.

World Meteorological Organization (2002), Standardized verification system (SVS) for long-range forecasts (LRF), in *Manual on the Global Data-Processing and Forecasting System*, vol. 1, Attach. 11.8, 5 pp., Geneva, Switzerland.

P.-S. Chu, C. W. Norton, and T. A. Schroeder, Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, 2525 Correa Rd., Honolulu, HI 96822, USA. (chasewn@hawaii.edu)