Statistical Downscaling of Climate Models with Deep Learning

Yusuke Hatanaka[a] , Amila Indika[a] , Thomas Giambelluca[b,c] and Peter Sadowski[a]

[a] *Information and Computer Sciences, University of Hawai'i at Mānoa, Hawai'i, USA*

[b] *Geography and Environment, University of Hawai'i at Mānoa, Hawai'i, USA*

[c] *Water Resources Research Center, University of Hawai'i at Mānoa, Hawai'i, USA*

Corresponding author: Peter Sadowski, psadow@hawaii.edu

ABSTRACT: Understanding climate change requires understanding how global changes in the atmosphere impact the climate at regional scales. Modeling this relationship in a data-driven manner is known as statistical downscaling. However, the amount of historical data needed to train statistical climate downscaling models is limited, so the field has relied almost exclusively on linear models. In this work, we propose a deep-learning approach to statistical downscaling. By reframing the learning problem to model local climate as a function of orography, seasonality, and coarse resolution atmospheric variables, we can train a single neural network model that generalizes to locations that have no historical data. In experiments, we demonstrate our approach by downscaling monthly rainfall in the Hawaiian islands and show that our method reduces the RMSE by 8% compared to an approach using linear models. We also show that qualitatively, the approach results in more detailed downscaled rainfall maps.

SIGNIFICANCE STATEMENT: A new method is proposed for statistically downscaling climate models to high spatial resolutions. In experiments, this deep learning approach is shown to be more accurate in downscaling monthly rainfall in Hawai‘i than a traditional statistical method. The method could improve predictions of future rainfall and other quantities in a changing climate, especially for locations where complex terrain results in highly localized climate patterns.

1. Introduction

Climate change is typically modeled using general circulation models (GCMs) that simulate large-scale physical features of atmospheric circulation. The relatively coarse spatial discretization of GCMs does not allow accurate simulation of processes at smaller spatial scales, which requires the additional step of *downscaling*. Downscaling methods convert coarse-resolution GCM projections to finer-resolution projections (Lauer et al. 2013; Ashfaq et al. 2022; Feyissa et al. 2023; Rahman and Pekkat 2024; Brands 2022; Virgilio et al. 2022). There are two principal downscaling approaches — statistical and dynamical — but dynamical downscaling requires physics-based simulations, for which the computational cost to model orographically complex regions such as Hawai‘i accurately is a limiting factor (Schmith 2008). Thus, statistical downscaling is an essential tool for modeling climate change in regions such as Hawai‘i (Elison Timm et al. 2015; Elison Timm and Diaz 2009; Elison Timm et al. 2011; Sanfilippo et al. 2023).

The amount of historical data available in Hawai‘i for training limits the accuracy of statistical downscaling models (Grotch and MacCracken 1991). Statistical downscaling models relate coarse resolution atmospheric variables to local variables of interest, enabling them to project future weather and climate by predicting these local variables from the output of GCMs. They are data-driven and typically fit using historical atmospheric data from reanalysis data products and local data from direct observations, such as precipitation measurements at the surface. The most commonly used reanalysis data products go back to the 1940s, with the uncertainty of these data products increasing as one goes further into the past (Kistler et al. 2001). However, there exist few high-resolution historical data products. Hence, in Hawai‘i, statistical downscaling models typically use simple, linear, statistical models (Elison Timm et al. 2015; Elison Timm and Diaz 2009; Sanfilippo et al. 2023). Some work has explored more sophisticated machine learning models (Hatanaka 2022; Norton et al. 2011), but the amount of training data limits the advantage

of more complex models. Indeed, machine learning has seen more success in downscaling weather than climate (Hatanaka et al. 2023; Hart et al. 2020) because the higher temporal resolution of weather observation data results in larger training datasets. Climate downscaling models typically operate on monthly or seasonal averages, so there are only a few hundred months of historical data to train and evaluate the models.

This work addresses the issue of limited training data by using ideas from transfer learning to reframe the machine learning problem. Most statistical downscaling approaches treat each geographic location as having a unique climate with very little influence from nearby sites via site-specific models; this flexibility enables the modeling of microclimates but requires abundant historical data at each site. In this work, we propose a model that predicts local climate variables as a function of atmospheric and orographic features. We argue that the combination of atmospheric and orographic properties largely determines the local climate, such that the advantages of a general, site-agnostic model outweigh those of modeling each location independently. We call this approach Location-Agnostic Neural Downscaling (LAND).

In experiments, we use LAND to downscale reanalysis data to predict the monthly precipitation for the Hawaiian islands at 250 m resolution. Hawai‘i is an interesting test case for multiple reasons. First, the mountainous orography of the islands produces steep rainfall gradients with mean annual values ranging from 200 mm to over 10,000 mm per year within the state (Sanderson 1994; Giambelluca et al. 2013), resulting in hyperlocal rainfall patterns and microclimates across the islands. Rainfall is high near the summits of Kaua‘i, O‘ahu’s two mountain ranges, Moloka‘i, and West Maui and on the northeast-facing slopes of East Maui, and the Big-island, where trade winds force persistent uplift along windward mountain slopes. Second, a high-quality historical rainfall dataset is available from the Rainfall Atlas of Hawai‘i (RAH) (Giambelluca et al. 2013), which comprises the combination of observational data and gap-filled rainfall data from 1920 to 2012. Third, there are significant implications for the results of this work, as long-term water resource management for the Hawaiian islands depends on accurate estimates of future rainfall under a changing climate.

2. Background

Previous statistical downscaling approaches have relied almost entirely on linear models. Some typical examples are Sanfilippo et al. (2023) and Elison Timm et al. (2015), which use a combination of linear dimensionality reduction and linear regression to make seasonal rainfall predictions at more than 850 weather stations in Hawai'i. Dimensionality reduction can be performed by principal component analysis (PCA) to model seasonal rainfall patterns as a combination of linear latent factors. Linear regression is then used to predict these latent factors from coarse atmospheric variables. The dimensionality reduction step helps prevent overfitting and helps to smooth out anomalous measurements at individual sites. However, the result is that the seasonal rainfall at each site is a linear function of atmospheric features.

Aside from the modeling limitations of linear models, this approach has other significant limitations. First, the model cannot leverage data from sites with few observations; each site needs enough observations to build a separate site-specific model. Second, because the model is only trained to make predictions at specific sites for which training data is provided, it cannot be used to make predictions at new locations without an additional modeling step. One solution to this problem is to interpolate between the predictions of the downscaling model using bilinear interpolation or Kriging (Lucas et al. 2022; Frazier et al. 2016) (since Kriging is better known as a Gaussian Process (GP) model in the machine learning literature, we use that terminology here). A major problem with this approach is that using the output of the downscaling model within a GP model leads to inaccurate uncertainty estimates. Using a GP to interpolate the observations and the resulting gridded product for training the downscaling model has the same problem. The proposed method solves this problem by using a single model that is generalizes in both the temporal and spatial domains.

3. Methods

This section begins by describing the proposed method, LAND. The focus is on downscaling monthly rainfall, but the approach could be applied to other climate variables and time intervals. Second, we describe alternative statistical downscaling approaches that we compare against LAND in experiments. Finally, we describe the experiments in detail, including datasets, hyperparameter tuning, and evaluation metrics.

a. Location-Agnostic Neural Downscaling (LAND)

LAND consists of a single neural network model that predicts monthly rainfall at any location within a region from a combination of atmospheric, orographic, and seasonal features (Figure 1a). Importantly, LAND is *location agnostic* in that predictions can be made at any location with no explicit dependence on the latitude and longitude coordinates. This enables LAND to perform statistical downscaling at arbitrarily high resolutions, even though the model is trained on sparse observations (see Figure 1b).

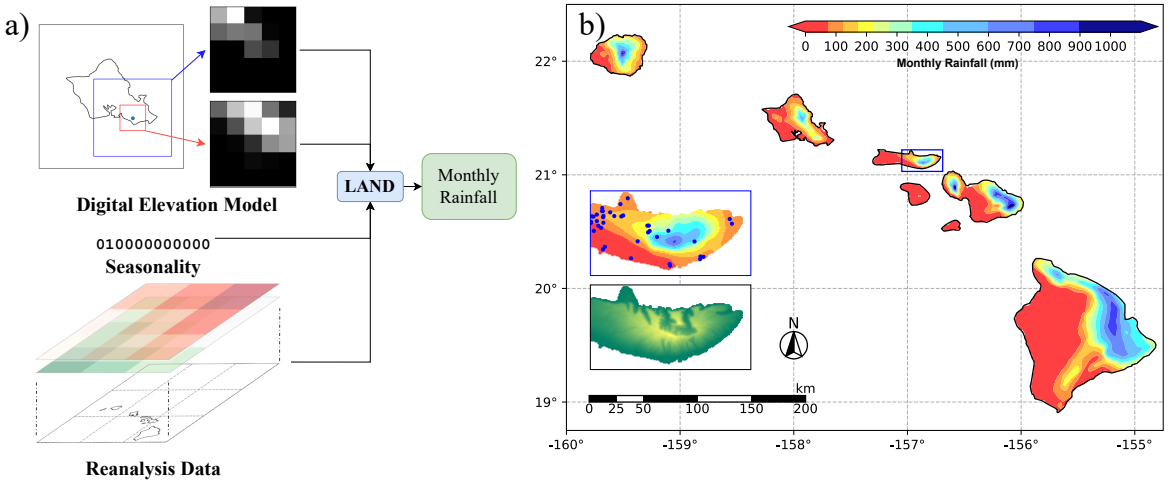


FIG. 1. (a) LAND makes predictions for monthly rainfall from local orographic and atmospheric features, represented as image-shaped inputs. The month is also provided as a one-hot vector to account for possible season-related variations in rainfall response to atmospheric patterns. (b) Predicted monthly rainfall (in mm) for January 2000. *Left-Middle*: Zoom of the blue rectangle over the island of Moloka‘i. The blue dots represent the locations of the weather stations in the training data. Weather stations on east Moloka‘i are sparsely distributed, but the model smoothly interpolates based on the orographic features. *Left-Bottom*: The elevation map on the zoomed region. The model accounts for the orographic information to predict high rainfall at the peak on the windward side of an island, which is a general pattern observed across Hawai‘i.

Orographic features are provided to LAND as a Digital Elevation Model (DEM). To capture both small-scale and mid-scale orographic features, a given location is described by a pair of orographic maps at different scales — each represented as 5-by-5 pixel images. Figure 2 illustrates this process for an example site on O‘ahu. Starting from a 250 m resolution DEM, a 20 km square region centered at the site is extracted and coarsened to 4 km resolution; we refer to this as the *local* DEM.

Similarly, a 60 km square region is extracted and coarsened to 12 km resolution; we refer to this as the *regional* DEM. The size of these regions was chosen to describe the position of a location relative to the mountains that heavily influence the weather in Hawai‘i. The resolution was chosen to balance the competing goals of including useful features and reducing the tendency to overfit.

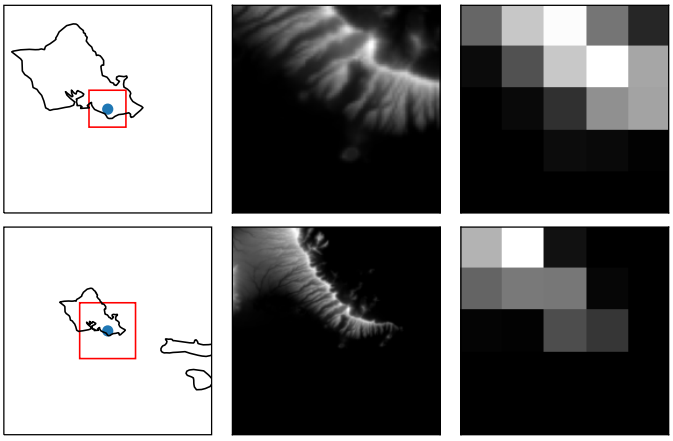


FIG. 2. Example of local (top) and regional (bottom) orographic features provided to LAND. Top-left: A 20x20 km region (red square) around a weather station (blue dot) is extracted from the DEM. Top-middle: The local DEM at full 250 m resolution. Top-right: The local DEM at coarse resolution; this is the input to the neural network model. The bottom shows the same processing for the 60x60 km regional DEM.

Atmospheric features are provided to LAND as a three-dimensional matrix, $\mathbb{R}^{c\times h\times w}$, where c represents the number of different atmospheric variables and h and w represent spatial dimensions. To capture seasonal effects, such as the path of the Sun during various times of the year, LAND also takes in the month as a one-hot-encoded vector. This results in a total of 206 model inputs and 250 thousand parameters. We employ careful hyperparameter tuning and ensemble ten networks with different random initializations to avoid overfitting and reduce the prediction variance. Appendix c provides the model structure and hyperparameter optimization details.

b. Comparison to Linear Statistical Downscaling Models

To demonstrate the advantages of the proposed method, we compare LAND to linear statistical downscaling approaches that do not use deep learning. The primary goal is to produce an accurate map of monthly rainfall, so the approach consists of two steps. First, linear regression is used to

make predictions at each weather station. Second, a GP model is used to interpolate between these sparse predictions to produce a complete map. Like LAND, this two-step approach can produce downscaled maps of arbitrary spatial resolution, but the model has a very different set of inductive biases. It also has a disadvantage in that the second model does not account for the uncertainty of the first model.

The first step is to fit a separate downscaling model for each weather station with historical rainfall data. The input variables to these models are the composite maps of reanalysis variables, and the target output is the monthly rainfall at the station. During preliminary experiments, we explored different types of models (linear regression vs. neural networks) and different variants of the composite maps (different map sizes). From these experiments, we concluded that linear models performed about the same as more complex neural networks (see Appendix d), with the additional advantage that they are less likely to overfit on sites with little training data. Thus, we use the simpler linear models in experimental comparisons with LAND. Because a separate model is fit for each weather station site, we refer to this method as site-specific linear regression (SSLR).

A GP is used to interpolate between weather stations to produce maps of predicted rainfall from predictions at individual stations. In our experiments, we obtain SSLR predictions for each test month and region, fit a Gaussian Process with a radial basis function (RBF) (see Appendix e), and conduct a thorough hyperparameter tuning using leave-one-out cross-validation (LOOCV) on the test set, regionally and monthly. This approach allows us to predict rainfall at the left-out station of LOOCV, ensuring a clean validation process. At each month, four GP models are independently fit to each of the four major regions: Kaua‘i, O‘ahu, Big Island, and Maui Nui, which consists of Maui, Moloka‘i, Lana‘i, and Kaho‘olawe. This approach has been used in other downscaling efforts for Hawai‘i Lucas et al. (2022). As a result, we obtain a predicted monthly rainfall at any location, achieving the same objective as LAND.

c. Datasets

1) Rainfall Observations

We retrieved the historical rainfall data from the Rainfall Atlas of Hawai‘i (RAH) (Giambelluca et al. 2013), which contains monthly rainfall from over 2000 rain gauges across Hawai‘i. This dataset comprises observational and spatially gap-filled rainfall data from 1920 to 2012. Gap-filling

methods have been applied to fill some missing data (Giambelluca et al. 2013) for SSLR but not for LAND, as described below.

2) Reanalysis Data

Reanalysis data approximates historical global climate variables and is produced by coarse-grained physics-based numerical models constrained by observations. In this work, we use data published by the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) (Kalnay et al. 1996). The retrieved data consist of monthly mean data for 16 variables representing temperature, humidity, atmospheric circulation, atmospheric stability, and moisture transport at different vertical levels on a 2.5° by 2.5° grid globally since 1948. The complete list of the 16 variables used as input to the model is available in Appendix a. To represent the atmospheric condition at each location, we created composite maps of the 16 variables for each month on a three-by-three grid over each location. This results in a three-dimensional matrix of dimensions (16, 3, 3) for each month, i.e., $c = 16, h = w = 3$. During model selection, we tested other spatial coverage as well, e.g., $(h, w) \in \{(5,5), (1,1), (2,3)\}$, which is described in Appendix c.

d. Evaluation

All experiments used data from 1948 to 1999 for model training and hyperparameter optimization, while data from 2000 to 2012 was reserved for evaluation. Furthermore, to test the models' ability to generalize to locations never seen in the training set, k-fold cross-validation (CV) is performed with respect to weather stations. Thus, models are trained on 1948–1999 data from a subset of the stations and evaluated on 2000–2012 data at the other stations. In the experiments with LAND, a 10-fold CV was performed, requiring ten different models to be trained. The SSLR models only needed to be trained once, and then the GP model was evaluated with leave-one-out cross-validation (LOOCV); thus, SSLR+GP is given a slight advantage over LAND in the experiments.

The training data for LAND consists only of observational rainfall data, while SSLR uses a combination of observational and gap-filled data. This resulted in 355,632 monthly rainfall values from 1,796 weather stations for training LAND and 667,632 observations from 1,102 weather stations for SSLR. We included gap-filled data for training SSLR to compensate for the missing data. The test set consists of all observational data from 2000 to 2012 from any weather station;

no gap-filled data were used for evaluation. This results in 56,620 data points from 515 distinct weather stations over 13 years of record. For more detail, see Appendix b.

We note that this experimental framework is known as *perfect prognostic observational downscaling* and assumes that the projection from reanalysis to rainfall is “perfect" and transferable to GCMs (Rampal et al. 2024). In practice, an ensemble of GCM projections could be used to model future climate uncertainty, and the downscaling models would be applied to each ensemble member.

e. Metrics

The models were evaluated using a variety of metrics. Let $Y = \{y_1, y_2, ..., y_n\}$ and $\hat{Y} = \{\hat{y}_1, \hat{y}_2, ..., \hat{y}_n\}$ denote the observed values and predictions, respectively. We evaluate the model performance based on R^2, Mean Absolute Error (MAE), Median Absolute Deviation (MAD), and the Root Mean Square Error (RMSE) as defined in Table 1.

TABLE 1. Metrics used for evaluation.

Metric	Definition	Unit
R^2	$R^2 = 1 - \frac{RSS}{TSS}$	Unitless
MAE	$\frac{1}{n}\sum \lvert \hat{y}_i - y_i \rvert$	mm
MAD	$\text{median}\{\lvert y_i - \hat{y}_i \rvert\}_{i \in N}$	mm
RMSE	$\sqrt{\frac{1}{n}\sum (y_i - \hat{y}_i)^2}$	mm

We also report the normalized version of some of the statistics. Values are divided by the mean of observations, and we denote those unitless quantities with $\widehat{\ }$ or prefix ‘r.’ For example,

$$\widehat{\text{RMSE}} = \text{rRMSE} = \frac{\text{RMSE}}{\overline{Y}}$$

4. Results

This section compares LAND and the combination of traditional approaches quantitatively and qualitatively. We demonstrate that LAND increases performance in terms of quantitative metrics

TABLE 2. Comparison of Error Metrics

Metric	All		Kaua‘i		O‘ahu		Maui Nui		Big Island	
	LR+GP	LAND	LR+GP	LAND	LR+GP	LAND	LR+GP	LAND	LR+GP	LAND
R^2 ↑	0.55	0.62	0.46	0.57	0.55	0.59	0.51	0.60	0.62	0.65
MAE ↓	58.98	52.41	74.59	68.97	60.41	52.36	49.35	44.14	63.36	56.00
$\widehat{\textbf{MAE}}$ ↓	0.54	0.48	0.52	0.48	0.51	0.44	0.61	0.54	0.51	0.45
MAD ↓	34.87	28.52	45.05	39.57	40.59	30.89	24.67	21.57	40.35	30.90
$\widehat{\textbf{MAD}}$ ↓	0.32	0.26	0.31	0.27	0.34	0.26	0.30	0.27	0.32	0.25
RMSE ↓	102.97	95.16	131.96	117.92	90.19	85.50	103.12	92.95	99.55	95.61
$\widehat{\textbf{RMSE}}$ ↓	0.94	0.86	0.92	0.82	0.77	0.73	1.27	1.14	0.79	0.76
N	56,620		6,836		14,953		20,032		14,799	

and results in qualitatively more realistic rainfall predictions that capture the extreme rainfall gradients found in Hawai‘i. We show that this improvement comes from including orographic information and transferring knowledge between locations.

a. Accuracy

The primary quantitative comparison is the RMSE on the test set of observations from 515 weather stations for months between January 2000 and December 2012. Table 2 shows a variety of performance metrics averaged across all stations, as well as averages for each island. Overall, LAND achieves a 12% increase in R^2 and a 7.5% decrease in RMSE compared to LR+GP. The error distributions are examined in more detail in Appendix g.

b. Qualitative Comparison

In this section we examine the qualitative properties of the predicted rainfall maps, which can provide further intuition for how LAND uses geographic features to improve performance. Rainfall predictions were plotted on a rasterized grid and compared to historical rainfall maps published in the Rainfall Atlas of Hawai‘i (RAH) (Giambelluca et al. 1986). The historical rainfall maps from the RAH do not represent direct observations at every location but were hand-drawn by climate scientists who combined weather station data and their expert understanding of local rainfall systems to produce best estimates. Typical examples of annual and monthly maps are shown in Figures 3 and 4, respectively.

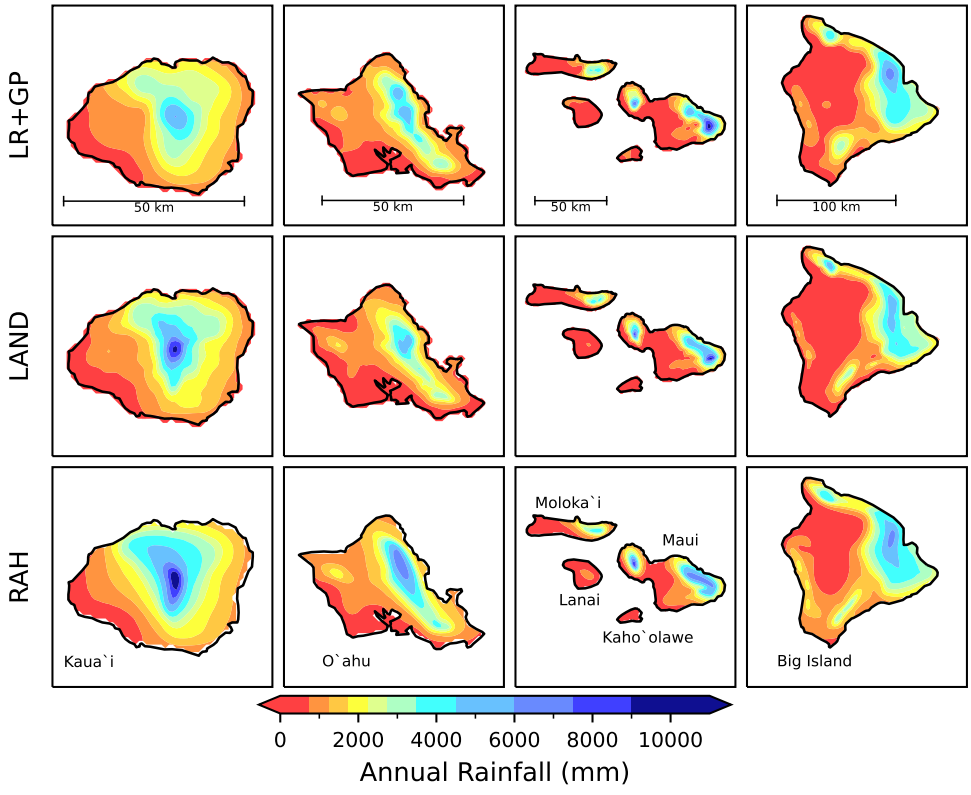


FIG. 3. Mean annual rainfall maps for LR+GP, LAND, and manually-drawn isohyets (RAH) (Giambelluca et al. 1986).

We observe that LAND predictions align better with these hand-drawn rainfall maps (Figure 3). For example, LAND and the RAH both show an elongated rainfall pattern following the mountain slope on the southeast side of the Big Island, while the LR+GP prediction appears as an isotropic spot. This makes sense since the GP kernel uses an isotropic RBF kernel to interpolate between stations while LAND uses the local orography, enabling LAND to capture highly localized rainfall patterns that occur against mountain slopes with a precision that is impossible with LR+GP. Another example is observed on Lana‘i, where LAND and RAH agree on rainfall max in the center of the island, while LR+GP predicts rainfall max on the north side of the island.

When we examine monthly rainfall maps (Figure 4), these qualitative differences are even more apparent. The rainfall maps from LR+GP often appear mottled, especially in the wet season (Jan., Mar., Nov.), due to the GP's interpolation between locations with highly variable rainfall. LAND is better suited for smoothly interpolating the inland area where weather station observations are

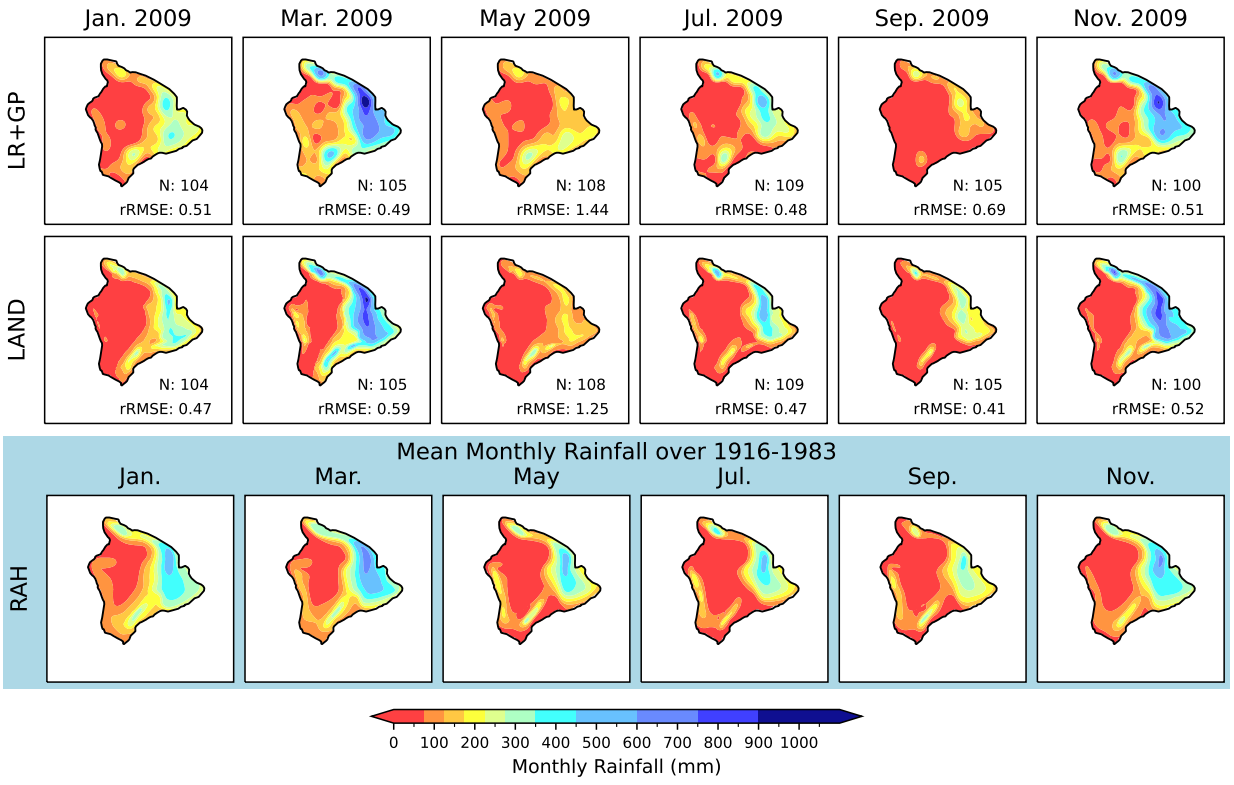


FIG. 4. Predicted rainfall maps from LR+GP (top row) and LAND (middle row) for six months in 2009. rRMSE and the number of stations (N) used to calculate the rRMSE are shown for each month. The bottom row shows the mean monthly maps from RAH over 1916-1983 for comparison (Giambelluca et al. 1986).

sparse (see Figure 6b). In Appendix f, we further compare LR+GP and LAND in capturing topographical effects.

c. Spatial Error Distribution

We mapped the spatial distribution of the prediction errors to better understand the advantages of LAND. RMSE and rRMSE were calculated at each weather station over the test period. Then, a GP was used to interpolate and visualize the error distribution across the state. Figures 5a and 5b show how RMSE and rRMSE are distributed, respectively. The RMSE pattern matches the general rainfall pattern of Hawai‘i, such that RMSE is high wherever rainfall is high. However, rRMSE tends to be high on the leeward side and inland area of the Big Island, where rainfall is lower (and hence the denominator is smaller).

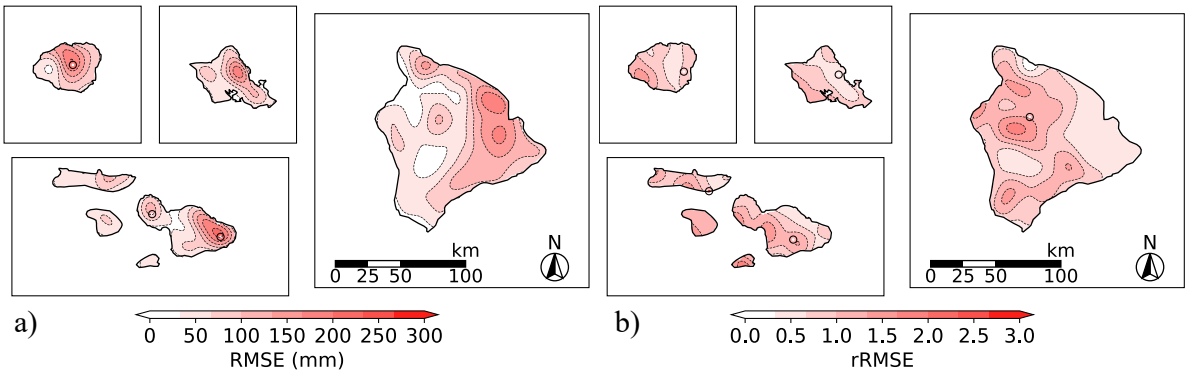


FIG. 5. Gaussian process interpolation of error. Black solid circles indicate weather stations where statistics are outliers; hence, they are excluded from interpolation. (a) Spatial distribution of RMSE. (b) Spatial distribution of rRMSE.

We also examined the spatial pattern of rRMSE improvement (Figure 6a). This quantity is defined as the difference between rRMSE values calculated on the predictions from LAND and LR+GP, where positive values indicate improvement and negative values indicate underperformance of the LAND approach. Again, this was computed for each weather station, and then a GP was used to interpolate for visualization.

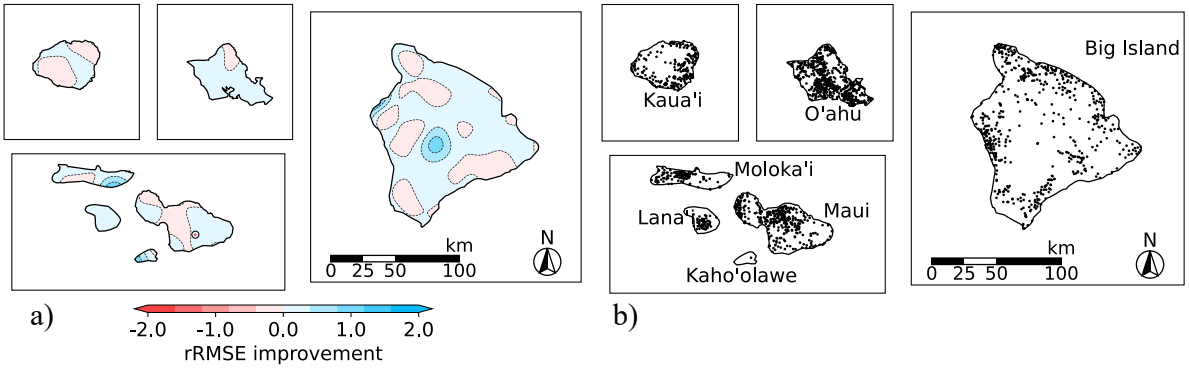


FIG. 6. (a) The difference in rRMSE (unitless) at weather stations, interpolated by GP for visualization. Positive values (blue shading) show where LAND improves upon LR+GP method (in terms of rRMSE). (b) Locations of weather stations used for training. LAND shows the largest improvements in regions with few weather stations, e.g., east Moloka‘i, central Big Island, and Kaho‘olawe.

Figures 6a and 6b show that LAND significantly improves in regions with few weather stations. For example, east Moloka‘i, central Big Island, and Kaho‘olawe, regions with few weather stations, are areas where LAND exhibits the most prominent performance improvements. This makes sense because LAND can leverage data from similar locations with similar orography to make predictions, while the LR+GP method ignores the local orography and relies on distant weather stations.

Additional evidence that LAND is making use of the DEM can be seen by examining the hidden activations of the neural network model. Figures 7a and 7b show how each pixel's DEM representation is clustered by the K-means algorithm before and after the input DEM is transformed by the initial layers of LAND. The clusters in Figure 7a do not correlate with rainfall or climate patterns, as all of Kaua‘i, O‘ahu, Moloka‘i, and Kaho‘olawe are clustered together. However, the neural network activation clusters in Figure 7b correspond to known climate regions, such as O'ahu's windward and leeward sides. In fact, the NW-facing and SE-facing regions of multiple islands are each clustered together (vertical and horizontal hatches, respectively), indicating that LAND recognizes similarities between these regions. Thus, LAND has learned a *climate embedding* that maps a raw DEM to a semantically meaningful feature space. This allows the model to transfer patterns learned at one location to different locations on other islands.

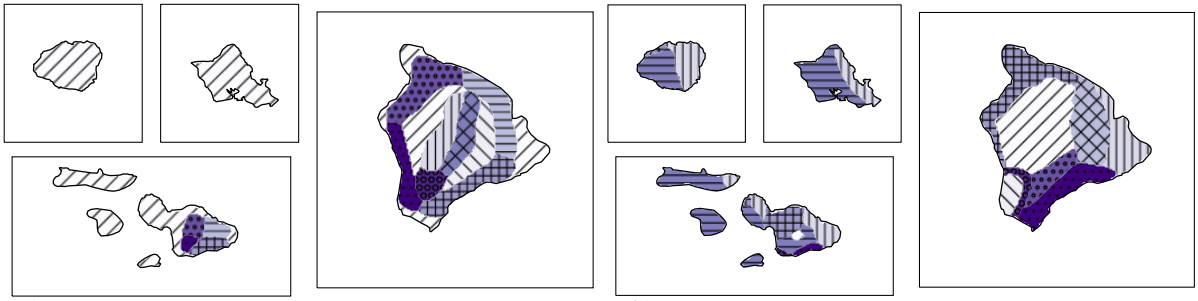


FIG. 7. K-means clustering of locations by (a) raw DEM features, and (b) learned embeddings in LAND (neural network activations after two dense layers are applied to the DEM). Note that the absolute rearrangement of the classes specified by colors and hatches between (a) and (b) is irrelevant in this context.

d. Case studies on selected stations

This section focuses on a time series of predictions on six stations across Hawai‘i, chosen based on their rainfall and geographical characteristics. Those stations include Kalaeloa Airport (dry,

leeward), Pu‘u O Hoku Ranch (remote from other weather stations), Big Bog (high elevation, wet), Kahuna Falls (wet, windward), and Kulani Mauka (high elevation, dry, remote). The locations are shown in Figure 8 with the map of LAND-predicted mean annual rainfall over 1948-1999 (the training period) to show the general rainfall characteristics of each location. The prediction error for each station is reported in Table 3, and Figure 9 shows the observed vs. predicted rainfall over time. LAND performs much better on the Big Bog site, decreasing the RMSE from 491 to 344 mm. This is explained by Big Bog being in an area with sparse station coverage (so LP+GP performs especially poorly) and Big Bog having high average rainfall (so RMSE is high in general). LAND shows a weaker performance advantage at the other sites, but the advantage is consistent across various climates and geographical characteristics.

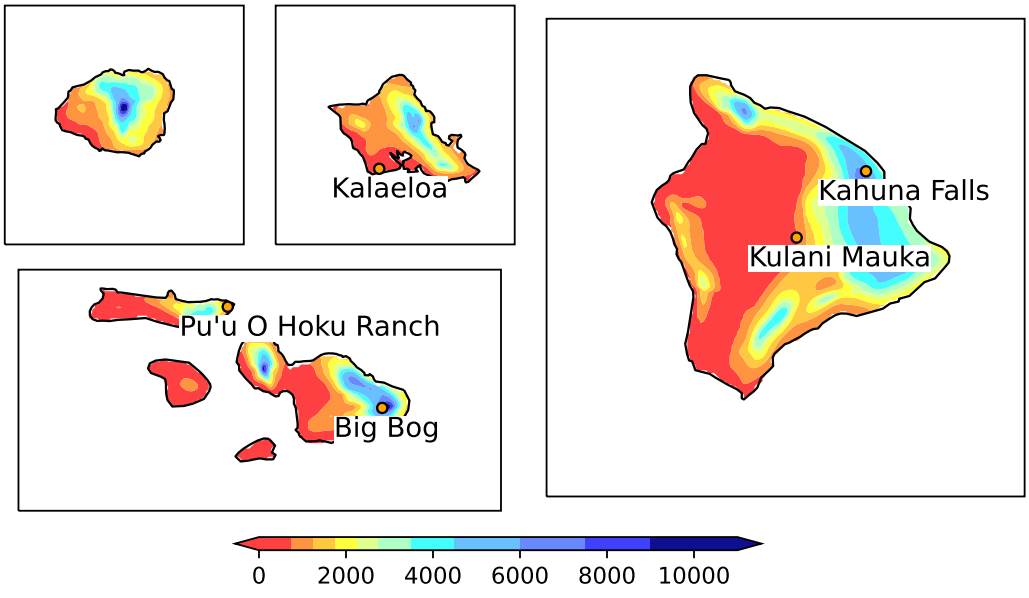


FIG. 8. Locations of weather stations for the case study, shown on the map of mean annual rainfall for 1948-1999 (mm), produced by our approach (LAND).

e. SSLR vs. LAND

So far, we have compared LAND against the LR+GP approach using cross-validation, where the evaluation was made on the left-out weather stations such that neither LAND nor LR had trained on the evaluation sites. Those results reflect the models' capacity to extrapolate the predictions to new locations without ever seeing data from the location. In this section, we compare the performance

TABLE 3. The performance of both methods on selected weather stations.

Station	**Kalaeloa**		**Pu'u O Hoku**		**Big Bog**		**Kahuna Falls**		**Kulani Mauka**	
	LR+GP	LAND	LR+GP	LAND	LR+GP	LAND	LR+GP	LAND	LR+GP	LAND
R^2	0.39	0.47	0.27	0.30	-0.28	0.38	0.43	0.45	0.07	0.37
MAE	27.81	22.15	43.02	41.94	351.81	243.75	155.61	154.85	43.82	27.35
$\widehat{\textbf{MAE}}$	1.00	0.80	0.52	0.51	0.47	0.33	0.37	0.37	0.86	0.54
MAD	17.15	10.83	27.91	29.38	241.97	169.53	125.90	129.29	39.96	16.32
$\widehat{\textbf{MAD}}$	0.62	0.39	0.34	0.36	0.33	0.23	0.30	0.31	0.79	0.32
RMSE	39.60	36.90	62.81	61.38	491.47	343.98	203.90	201.29	51.69	42.34
$\widehat{\textbf{RMSE}}$	1.43	1.33	0.76	0.75	0.66	0.46	0.49	0.48	1.02	0.83
N	156		139		117		151		156	
Island	O'ahu		Moloka'i		Maui		Big Island		Big Island	
Features	Low rainfall		Remote		High rainfall		High rainfall		High elevation	
	Leeward		Complex terrain		High elevation		Windward		Inland	

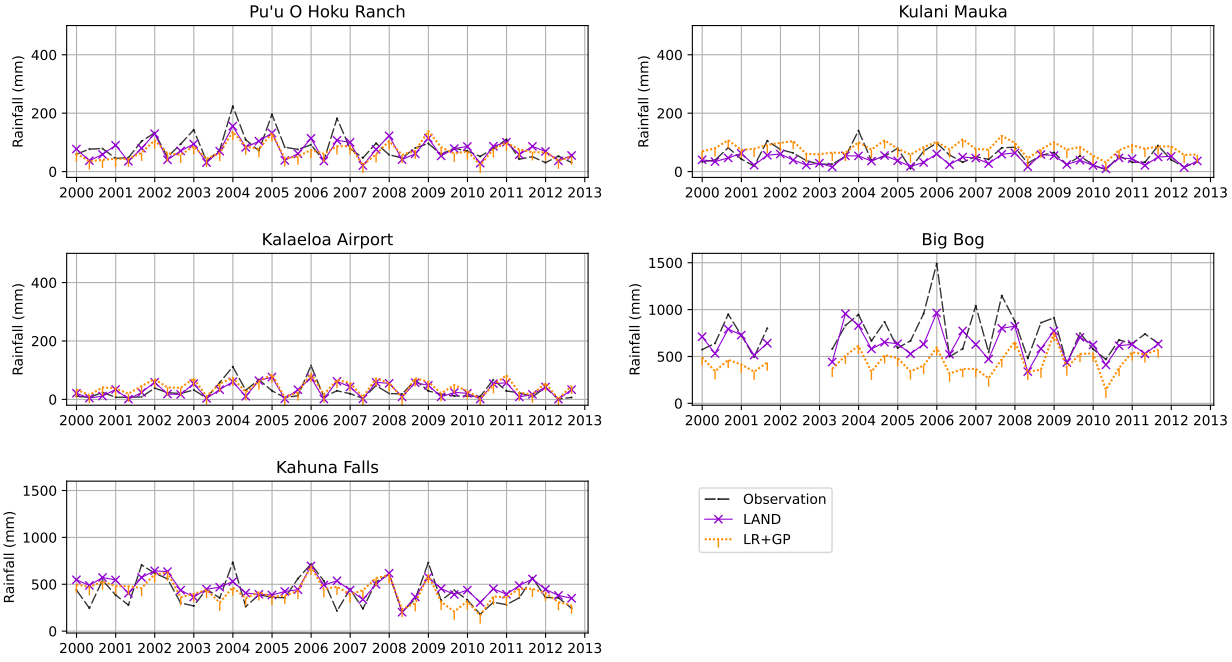


FIG. 9. Time series of observed vs. predicted monthly rainfall (in mm) aggregated into three-month bins for clarity.

between SSLR and LAND on locations *with* historical data, without any GP interpolation. That is, we evaluate performance on locations for which historical observation data exists, which allows us to compare the predictions of LAND directly against SSLRs.

TABLE 4. Comparison of Error on Locations with Historical Data Available

Metric	All		Kaua‘i		O‘ahu		Maui Nui		Big Island	
	SSLR	LAND	SSLR	LAND	SSLR	LAND	SSLR	LAND	SSLR	LAND
R^2	0.64	0.66	0.68	0.67	0.59	0.62	0.62	0.67	0.65	0.67
MAE	59.06	54.36	67.17	65.50	58.55	51.91	56.84	53.10	57.30	51.79
$\widehat{\textbf{MAE}}$	0.49	0.45	0.44	0.43	0.48	0.43	0.55	0.51	0.48	0.44
MAD	34.84	29.53	44.38	39.49	38.48	30.19	27.70	25.30	33.29	28.41
$\widehat{\textbf{MAD}}$	0.29	0.25	0.29	0.26	0.32	0.25	0.27	0.24	0.28	0.24
RMSE	97.84	94.34	107.13	109.36	87.95	84.55	104.95	97.74	93.44	90.69
$\widehat{\textbf{RMSE}}$	0.81	0.78	0.71	0.72	0.72	0.69	1.01	0.94	0.79	0.76
N	34,464		5,352		10,186		10,713		8,213	

In this experiment, models are again trained on data from 1948-1999, and predictions are made for years between 2000 and 2012. Prediction performance is evaluated on 289 weather stations for which data is available in both the training and test sets.

LAND outperforms SSLR on all metrics when aggregating over all locations. This provides strong evidence that patterns learned at one location can help improve predictions at other locations. This advantage is consistent for each individual island except Kaua‘i, where LAND gets slightly higher RMSE, rRMSE, and R^2. Kaua‘i has a history of extreme rainfall events, and the fact that MAE and MAD are smaller than SSLR indicates that LAND is underestimating rainfall for some large rainfall observations. This suggests that some idiosyncratic locations may always be modeled best by site-specific models. The LAND method could be extended to fall back on site-specific models for particular locations where abundant historical data is available.

5. Discussion

The experimental results support the hypothesis that LAND provides a performance increase over traditional statistical downscaling methods. It is worth emphasizing that this performance increase comes from multiple advantages. We discuss these advantages and how they come with significant limitations that are yet to be fully understood.

The primary advantage of LAND is that it effectively increases the amount of training data to learn from. With any approach that learns site-specific parameters (such as SSLRs), fitting the parameter is restricted by the availability of data from its exact location. However, since it doesn't matter which station training data for LAND comes from, the model can train on more training

data. This, in turn, allows data from newly installed weather stations to become part of training data immediately. In contrast, it would take new weather stations decades to collect enough historical rainfall data to fit site-specific parameters. The fact that the model trains on data from all locations also potentially acts as regularization. Data collected from weather stations can be influenced by factors not representative of the local rainfall pattern (e.g., instrumental/calibration error or measurement error due to environment, such as high wind and orientation of slope (Giambelluca et al. 1986)). In this case, the site-specific model cannot correct the bias, and the parameters will overfit the artifact, whereas predictions from LAND leave room for regularization via other training data with similar DEM features.

The second advantage is the ability to make predictions at any location, removing the need for a two-step modeling process. The two-step process suffers from a problem in which predictions from the site-specific models are regressed towards the mean, so the spatial interpolation model experiences a domain shift between training and prediction time. The results in Section e show that the SSLR models perform closer to LAND when no spatial interpolation is necessary.

The third advantage is that there is no need for gap-filling. Weather stations with many missing data must be gap-filled to fit site-specific parameters. However, this process is unnecessary for LAND as long as it has enough training data, collectively from any weather stations covering various orographic features across the study area, which is the case with Hawaiian islands (Figure 6b). This is especially helpful for historical climate datasets as most weather stations are installed and/or decommissioned during the dataset's timeline.

On the other hand, LAND has several limitations. The model assumes that the atmosphere's interaction with orography primarily determines the rainfall at each location. There is a persistent pattern in Hawai‘i, where regular trade winds bring much more rain to the windward sides of the islands than to the leeward sides. Our results show that LAND learns these relationships in Hawai‘i, but it is not clear whether such patterns will generalize to other regions. Thus, our results open avenues for future work.

A second limitation of this work is that the deterministic predictions consist of point estimates at each site. However, probabilistic predictions are of great interest for risk management. A straightforward way to obtain probabilistic forecasts from the LAND framework is to use a heteroskedastic output prediction layer, in which the neural network outputs the parameters of a known distribution

family at each pixel, for example, the mean and the standard deviation of a Gaussian distribution. However, this approach assumes the independence between grid cells and would thus be inappropriate for modeling climate risks such as floods. Other methods for statistical downscaling using more sophisticated machine learning models that explicitly model these joint distributions have recently been proposed (Hatanaka et al. 2023).

6. Conclusion

We have presented a deep learning approach to statistical downscaling for climate variables. Importantly, this is not simply a replacement of traditional models with neural networks but a reframing of the statistical downscaling problem in a way that leverages the ability of deep neural networks to generalize in high-dimensional data space. We demonstrate that the method outperforms the traditional statistical downscaling approach through experiments on downscaling monthly rainfall in Hawai‘i using reanalysis and a large historical dataset. Analysis shows that this method is particularly advantageous in scenarios where data is sparse relative to the spatial variability of the data. The limitations of the proposed method are discussed, and future work is needed to understand the full range of applications for which the method could be valuable.

Acknowledgments. Support for this work comes from NSF #OIA-2149133, NSF #2238375, and PI-CASC G21AC10381. Technical support and advanced computing resources from the University of Hawai'i Information Technology Services – Cyberinfrastructure, funded in part by the National Science Foundation CC* awards #2201428 and #2232862 is gratefully acknowledged.

Data availability statement. The historical rainfall data used in this study is publicly accessible through the Rainfall Atlas of Hawai'i (RAH) `www.hawaii.edu/climate-data-portal/rainfall-atlas/`. Reanalysis data for experiments is available via the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) `psl.noaa.gov/data/gridded/data.ncep.reanalysis.html`.

APPENDIX

Appendix A

a. Reanalysis Variables

For building downscaling models, we use the following 16 variables listed in Table A1. These variables are consistent with the work by Sanfilippo et al. (2023).

The air temperature difference and potential temperature difference are calculated by first retrieving the values at different levels independently and subtracting one from another. Moisture transport is the element-wise multiplication of specific humidity and wind component at each grid point. For example, meridional moisture transport at 700 hPa is calculated by multiplying specific humidity at 700 hPa and v-wind at the same level at each grid.

b. Data Split

We use data from 1948 to 1999 (training dataset) for model training, and data from 2000 to 2012 (test dataset) for model evaluation. Since SSLRs require long historical rainfall data for training, we included gap-filled data for SSLRs, which resulted in 1,102 SSLRs with the total of 667,632 training data. On the other hand, LAND does not require each weather station to have dense history of observational data. We excluded gap-filled data for training LAND in order to exclude additional uncertainty from the gap-filing, which resulted in 1,796 unique weather stations and the total of 335,632 observational data.

TABLE A1. Climate features used as input to the downscaling models.

Feature
Geopotential Height at 500hPa
Geopotential Height at 1000hPa
Air temperature difference (1000hPa and 500hPa)
Surface air temperature at 2m
Zonal moisture transport at 700hPa
Zonal moisture transport at 925hPa
Meridional moisture transport 700hPa
Meridional moisture transport 925hPa
Omega
Specific humidity at 700hPa
Specific humidity at 925hPa
Precipitable water
Potential temperature difference (850hPa and 1000hPa)
Potential temperature difference (500hPa and 1000hPa)
Sea level pressure
Skin temperature

c. LAND Model Details

1) MODEL STRUCTURE

LAND is a neural network that uses local and regional DEMs, month, and reanalysis data as input variables (Figure A1b). The atmospheric variables are provided as a composite map of shape height h, width w, and channels c, centered at the nearest pixel. For example, Figure A1a illustrates how all stations in Cell 1 receive the composite map in the blue square, while all stations in Cell 6 receive the composite map in the yellow square; in both cases, the shape is $\mathbb{R}^{c\times h\times w} = \mathbb{R}^{16\times 3\times 3}$.

During hyperparameter optimization, we explored various kernel shapes: $(h,w) \in \{(1,1),(2,3),(3,3),(5,5)\}$. With $(h,w) = (2,3)$, all stations receive Cells 1 to 6 in Figure A1a, rather than centering at the nearest neighbor. Similarly, when $(h,w) = (5,5)$, we used the entire composite map within Figure A1a. Kernel shape $(3,3)$ was found to work best.

2) DATA PREPROCESSING AND TRAINING SETUP

DEM and rainfall data were divided by factors of 4,000 and 100, respectively, to scale the values so that the variance is approximately one. Each variable of the reanalysis composite maps

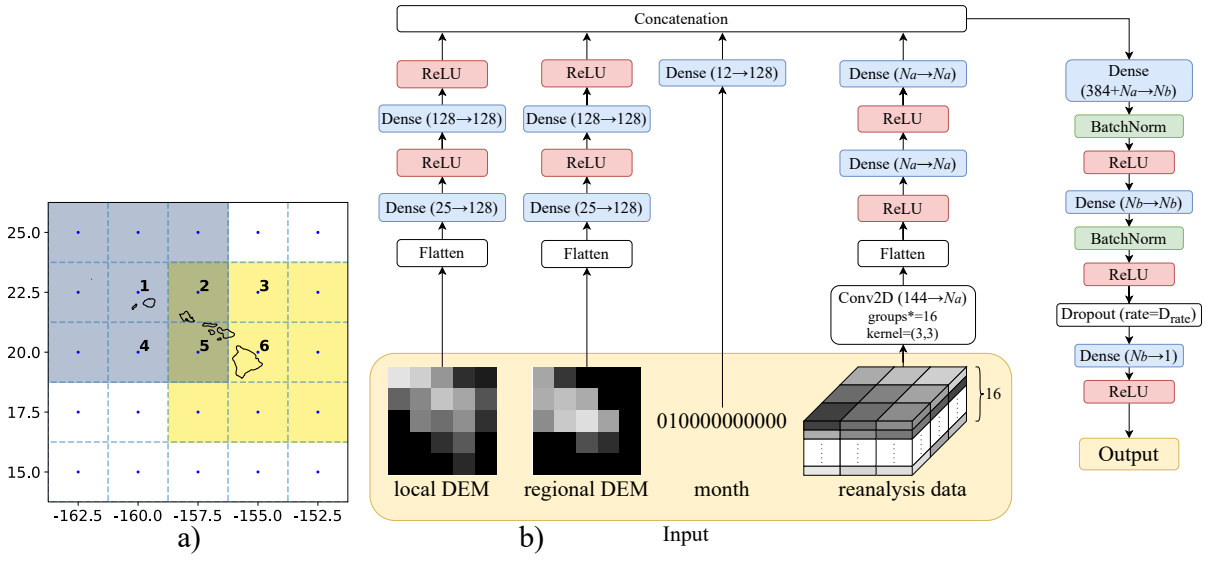


FIG. A1. (a) The reanalysis input to LAND is a 3x3 composite map. For example, any site in Cell 1 receives the composite map over the blue square, and any site in Cell 6 receives the composite map over the yellow square. (b) Model structure diagram. N_a and N_b represent the number of neurons. D_{rate} is the dropout rate. These are tunable hyperparameters of the model. The convolution layer is locally connected so that each output channel is only connected to a single input channel (*groups* = 16).

was normalized by subtracting the mean and dividing by the standard deviation over all pixel values in the data from 1948 to 2014. We used the Adam optimizer with decoupled weight decay (Loshchilov and Hutter 2019) to minimize the Mean Squared Error (MSE). Cosine annealing with warm restarts (Loshchilov and Hutter 2017) was used, where we fixed T_0 = 10, T_{mult}=2, and the maximum total number of epochs to be 150, which means the training completes just before the fourth warm restart. Randomly chosen 20% of the training data was set aside as a validation set. After the training, the model weights at the epoch with the lowest validation error were retrieved. Any data with rainfall below 0.1 mm or above 2,500 mm were removed from training data, as it was found that values outside the range could indicate unrealistic and erroneous records.

3) HYPERPARAMETER OPTIMIZATION AND MODEL STRUCTURE SEARCH

For hyperparameter optimization, we further split the training dataset into two subsets: data before 1989 (inclusive) for training and data after 1989 (exclusive) for evaluation. To avoid confusion, we redefine those subsets as *training dataset* and *validation dataset* (italicized), respectively.

We explored some model choices and hyperparameters by hand and others with an automated hyperparameter search. First, a selection is made from the Table A2, after which we run an automated hyperparameter search for the Table A3 using Tree-structured Parzen Estimator (TPE) implemented in the `optuna` package for 200 iterations (Akiba et al. 2019), where the model is trained on *training dataset*. We then pick the model that resulted in the lowest MSE on the *validation dataset*. Every time we modified a hyperparameter from Table A2, we repeated the optimization with `optuna` for the hyperparameters in Table A3. Note that this search was not exhaustive, but in total, over 3,000 different hyperparameter combinations were explored.

TABLE A2. Set of hyperparameters and model choices tuned by hand. The search is not exhaustive.

Hyperparameter / Model Choice	**Range**	**Best**
The last activation function	{softplus, ReLU}	ReLU
Optimizer	{Adam, AdamW}	AdamW
Composite map kernel for Reanalysis	{1x1, 2x3, 3x3, 5x5}	3x3
Month feature embedding	{positional embedding, one-hot}	one-hot
DEM branch first layer	{Flatten, Conv2D}	Flatten
Activation of Reanalysis branch	{ReLU, None}	None

TABLE A3. Hyperparameters tuned with `optuna`. `float` and `int` indicate the range explored, with the first and the second values indicating the lower and the upper bound of the searched values, respectively. The third value is the step size, or if 'log,' then it means the sample was taken uniformly in the log domain.

Hyperparameter	**Range**	**Best**
N_a	{256, 512}	512
N_b	{256, 512, 768, 1024}	1024
D_{rate}	`float`(0, 0.5, 0.05)	0.45
batch size	`int`(256, 1024, log)	314
initial lr	`float`(5×10^{-4}, 0.01, log)	1.17×10^{-3}
weight decay	`float`(5×10^{-4}, 0.01, log)	6.45×10^{-3}

d. Preliminary Experiment

This preliminary experiment aimed to 1) determine the best variant of the composite map as the input to the site-specific models, 2) examine the importance of data availability, and 3) compare the performances of SSLRs, SSNNs, and LAND on the selected sites. For this experiment, we focused on the subset of the stations where observational data was almost consistently available

between 1948 and 2012, and any stations with more than 5% of missing data during those years were filtered out. This resulted in 24 stations across Hawai'i.

SSLRs and SSNNs were trained on each of the 24 stations and made predictions for 2000-2012. We varied the amount of training data to examine the effect of the number of training samples. For $y \in \{1948, 1955, 1960, 1965, ..., 1995\}$, we train the model using data from years between y and 1999. For example, if $y = 1948$, the model trains on all the possible rainfall data for the station, and if $y = 1995$, the model trains only on five years' worth of data. LAND was also trained with variable length training data in the same manner, except it receives all available observational data from any station, not limited to the 24 stations.

The input variables to the SSLRs and SSNNs are reanalysis data. The two variants of the composite map tested were single-cell (*1×1*) or grid-cells (*2×3*). With the single-cell variant, each station receives the composite map only on the exact cell to which the station belongs. With the grid-cells variant, the composite map is consistent across all stations for every month and is the ones that cover the Hawaiian islands (Cells 1 to 6 in Figure A1a). This results in 16 and 96 input variables for the single-cell and grid-cell variants, respectively.

Figure A2 shows the result. In general, more training data leads to better performances across all models. All three site-specific model variants (SSNN-1x1, SSNN-2x3, and SSLR-1x1) perform similarly with no clear winner. In Hatanaka (2022), SSNN-2x3 achieved a better performance than SSLR-1x1, which was not observed in this experiment — this is because hyperparameter optimization was done to obtain the optimal hyperparameters for every SSNN, which necessitates as many hyperparameter-optimization iterations as the number of the site-specific models. It would, therefore, be implausible to scale it up to run hyperparameter optimization at every one of over a thousand SSNNs and maintain the models as would be necessary for making predictions as the first of the two-step (LR+GP) approach. For this reason, SSNN was rejected, and instead, SSLR-1x1 was used for the two-step approach, as it is computationally efficient while performing similarly to other site-specific model variants.

e. Gaussian Process

Gaussian process is a kernel-based method that allows the interpolation of samples in a given coordinate system. A kernel computes a covariance matrix of a multivariate Gaussian distribution,

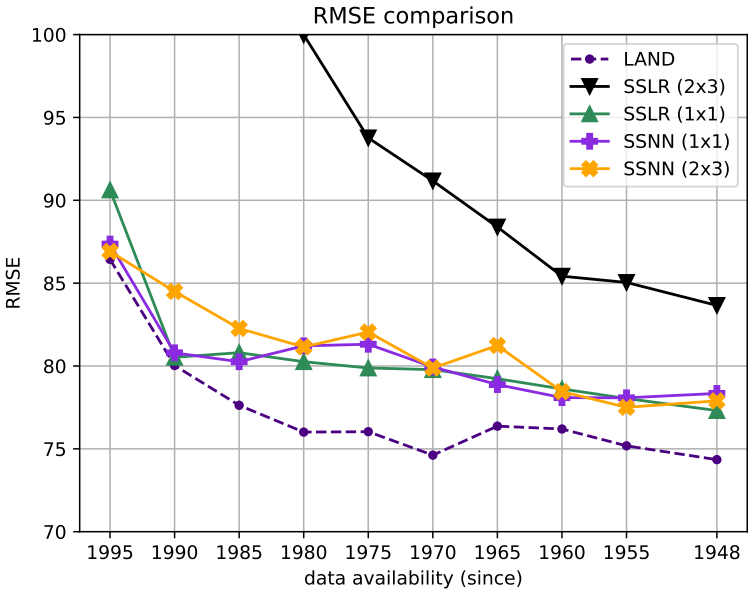


FIG. A2. Performance of LAND and site-specific models with different amounts of training data. The x-axis is the earliest year used for training—older dates correspond to more training data and improved performance. LAND performs similarly to the simple site-specific models when there is little historical data, but outperforms them when there is more training data.

after which new samples can be drawn from the posterior distribution under observation. Given coordinates $X = \{x_1, x_2, ..., x_n\}$, a kernel K computes the covariance matrix Σ between every pair of coordinates

$$\Sigma_{i,j} = \alpha K(x_i, x_j) + gI$$

where I is the identity matrix, and α and g are hyperparameters of the model, controlling the scale of the covariance and the independent homoskedastic noise at each observation, respectively. For the kernel function, we use one of the most commonly used kernels, radial-based kernel (also known as RBF kernel), as defined below

$$K(x_i, x_j) = \exp\left(-\frac{1}{2}(x_i - x_j)^T \Theta^{-2}(x_i - x_j)\right) \tag{A1}$$

where Θ, length-scale, is another hyperparameter controlling how strongly two points are correlated as a function of distance. As shown so far, three new hyperparameters are introduced: α, g, and

Θ. Despite slight differences in formulation, these three are analogous to *sill*, *nugget*, and *range* in Kriging (Christianson et al. 2023). GPyTorch is a Python package that implements the Gaussian process and utilizes gradient descent to optimize these hyperparameters on the likelihood of data under the hyperparameters (Gardner et al. 2018).

For the preprocessing of the target variable, the long-term mean and the standard deviation of the rainfall values were calculated using data from 1948 to 1999 to standardize the target variable. Another alternative for preprocessing is to use log-transformation

$$\hat{y} = \log(y + 1)$$

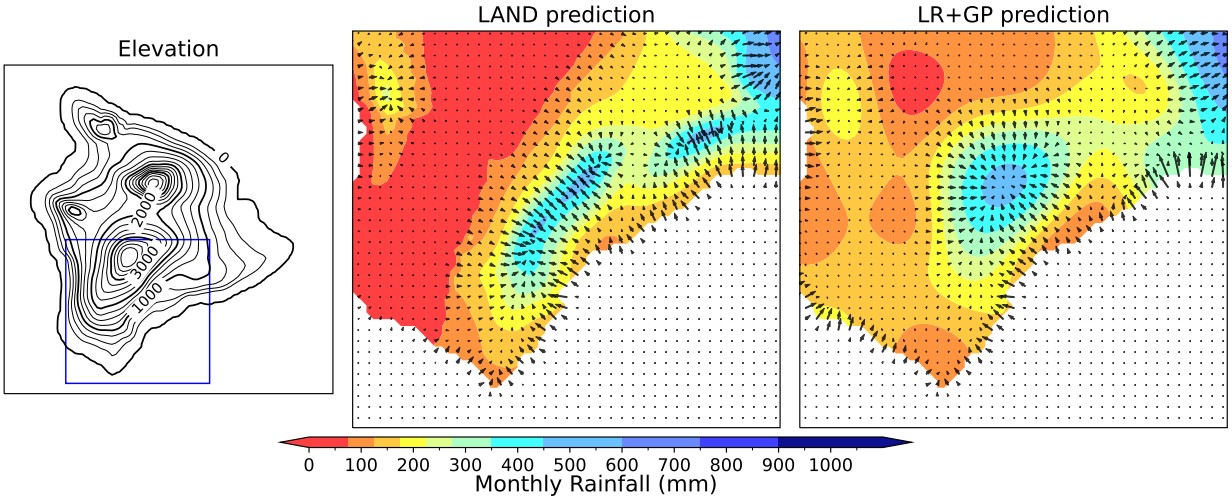
We tested both methods on a train-validation split (1948-1989 vs. 1990-1999) and found that standardization yields a better result on the validation set regarding MSE for the baseline.

Zero clipping is applied both before and after fitting the Gaussian process model. In other words, predictions from SSLRs are clipped to non-negative values before being fed to the Gaussian process model. Once the Gaussian process interpolates the predictions to a new location, negative values are clipped to zero again. This gives additional advantages to the LR+GP framework.

f. Details on Qualitative Evaluation of Rainfall Gradient

Intuitively, LAND predictions appear more accurate than LR+GP predictions because they more closely follow the orography. For example, Figure A3 shows monthly rainfall on the South of the Big Island for March 2009. We observe a region of high rainfall that appears very isotropic (round) in the LR+GP prediction, which conflicts with the local orography. This can be explained by the GP having no local orographic information. In contrast, LAND "sees" the local orography and it has learned that rainfall in Hawai‘i is often highly-localized along steep slopes, and this is reflected in its predictions.

We quantified this phenomenon by computing the cosine similarity of the gradient fields between the orography and rainfall predictions. First, gradient fields were calculated for predictions from LAND and LR+GP over the entire Hawai‘i for each month, as well as the DEM. Next we calculated the absolute value of the cosine similarity of the gradient fields between (1) LAND and DEM, and between (2) LR+GP and DEM, at each pixel. Higher absolute cosine similarity indicates that the direction of the rainfall gradient either aligns or opposes the direction of the orography gradient.



535 FIG. A3. Gradient field of LAND and LR+GP rainfall prediction. LAND predicts rainfall distribution that is
536 consistent with the local orography, whereas LR+GP tends to predict isotropic concentration of rainfall.

526 In other words, it indicates that the rainfall prediction increases or decreases in the same direction
527 as the slope of the land. Though this is not always the case with the actual rainfall pattern, we
528 expect rainfall to be distributed according to the local orography, as rainfall in Hawai‘i is greatly
529 influenced by local orography (Nullet and McGranaghan 1988). The absolute cosine similarity was
530 calculated at each grid point, and the monthly mean was calculated over all pixels above the sea
531 level. This was repeated over all months in the test period, after which the mean and the standard
532 deviation were calculated to be 0.731 ± 0.01 for LAND, and 0.689 ± 0.01 for LR+GP. We found
533 that the absolute cosine similarity is significantly higher for LAND, which indicates alignment
534 between rainfall and the orography in their gradients.

537 *g. Error Distribution*

538 This section discusses the models' performance in terms of R^2. This metric compares the model's
539 performance against a simple mean predictor, in which case $R^2 = 0$, while $R^2 = 1$ indicates a perfect
540 model with no error. R^2 is robust to absolute measurement unit (as opposed to $RMSE$) or division
541 by small values for the dry area (as opposed to $rRMSE$). Figure A4a shows the distribution of R^2
542 per weather station. The violin plots were created by calculating R^2 values at each weather station
543 and then plotting the distribution of R^2 along with the box plot. Black dots represent outliers,
544 and the plot is clipped at -1.5 for visualization. Any stations with less than 30 data points were

excluded to get reliable statistics. Figure A4b was created similarly, except the R^2 calculation was done every month.

Regarding Figure A4a, LAND achieves a smaller error dispersion over R^2 compared to the baseline with improved mean R^2 values. The distribution is more concentrated at higher R^2 values with stable improvement or comparable median to the baseline. The same trend is also observed in Figure A4b. For Kaua'i, the dispersion is more prominent, and the mean R^2 is worse than the baseline, but the improvement in the median R^2 value is prominent. This is because a few outliers resulted in poor R^2 values while the majority of other months resulted in improvement.

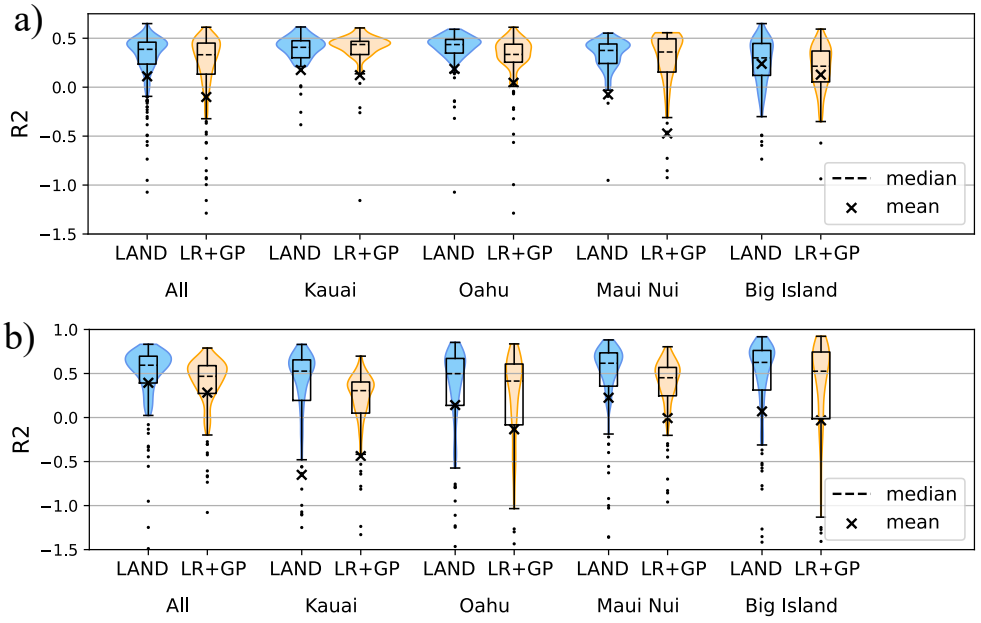


FIG. A4. (a) Distribution of R^2 across sites (b) Distribution of R^2 across months.

References

Akiba, T., S. Sano, T. Yanase, T. Ohta, and M. Koyama, 2019: Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, NY, USA, 2623–2631, KDD '19, https://doi.org/10.1145/3292500.3330701, URL https://doi.org/10.1145/3292500.3330701.

Ashfaq, M., D. Rastogi, J. Kitson, M. A. Abid, and S.-C. Kao, 2022: Evaluation of CMIP6 GCNs over the CONUS for downscaling studies. *Journal of Geophisical Research: Atmospheres*, **127**.

Brands, S., 2022: A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the northern hemisphere mid-to-high latitudes. *Geoscientific Model Development*, **15**, 1375–1311.

Christianson, R., R. Pollyea, and R. Gramacy, 2023: Traditional kriging versus modern Gaussian processes for large-scale mining data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **16**, https://doi.org/10.1002/sam.11635.

Elison Timm, O., H. Diaz, T. Giambelluca, and M. Takahashi, 2011: Projection of changes in the frequency of heavy rain events over Hawaii based on leading pacific climate modes. *Journal of Geophysical Research: Atmospheres*, **116 (D4)**.

Elison Timm, O., and H. F. Diaz, 2009: Synoptic-statistical approach to regional downscaling of IPCC twenty-first-century climate projections: seasonal rainfall over the Hawaiian islands. *Journal of Climate*, **22 (16)**, 4261–4280.

Elison Timm, O., T. W. Giambelluca, and H. F. Diaz, 2015: Statistical downscaling of rainfall changes in Hawai'i based on the CMIP5 global model projections. *Journal of geophysical research: Atmoshperes*, **120**, 92–112.

Feyissa, T. A., T. A. Demissie, F. Saathoff, and A. Gebissa, 2023: Evaluation of general circulation models CMIP6 performance and future climate change over the Omo river basin, Ethiopia. *Sustainability*, **15 (8)**.

Frazier, A. G., T. W. Giambelluca, H. F. Diaz, and H. L. Needham, 2016: Comparison of geostatistical approaches to spatially interpolate month-year rainfall for the Hawaiian islands. *International Journal of Climatology*, **36 (3)**, 1459–1470.

Gardner, J. R., G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, 2018: GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. *Advances in Neural Information Processing Systems*.

Giambelluca, T. W., Q. Chen, A. G. Frazer, J. P. Price, Y.-L. Chen, P.-S. Chu, J. K. Eischeid, and D. M. Delparte, 2013: Online Rainfall Atlas of Hawai'i. *Bulltein of the American Meteorological Society*, **94**, 313–316.

Giambelluca, T. W., M. A. Nullet, and T. A. Schroeder, 1986: Rainfall atlas of Hawai'i. State of Hawai'i, Department of Land and Natural Resources.

Grotch, S. L., and M. C. MacCracken, 1991: The use of general circulation models to predict regional climatic change. *Journal of climate*, **4 (3)**, 286–303.

Hart, K., P. Sadowski, and G. Torri, 2020: Nowcasting solar radiance over Oahu. *AI for Earth Sciences Workshop at NeurIPS*.

Hatanaka, Y., Y. Glaser, G. Galgon, G. Torri, and P. Sadowski, 2023: Diffusion models for high-resolution solar forecasts. URL https://arxiv.org/abs/2302.00170, 2302.00170.

Hatanaka, Y. M., 2022: Machine learning based statistical downscaling for rainfall on Hawaiian islands. M.S. thesis, Information and Computer Sciences, University of Hawai'i at Manoa.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, **77 (3)**, 437–472.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Sciety*, **82**, 247–268.

Lauer, A., C. Zhang, O. Elison-Timm, Y. Wang, and K. Hamilton, 2013: Downscaling of climate change in Hawaii region using CMIP5 results: On the choice of the forcing fields. *Journal of Climate*, **26**, https://doi.org/https://doi.org/10.1175/JCLI-D-13-00126.1.

Loshchilov, I., and F. Hutter, 2017: SGDR: stochastic gradient descent with warm restarts. *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, OpenReview.net, URL https://openreview.net/forum?id=Skq89Scxx.

Loshchilov, I., and F. Hutter, 2019: Decoupled weight decay regularization. *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, OpenReview.net, URL https://openreview.net/forum?id=Bkg6RiCqY7.

Lucas, M. P., R. J. Longman, T. W. Giambelluca, A. G. Frazier, J. Mclean, S. B. Cleveland, Y.-F. Huang, and J. Lee, 2022: Optimizing automated kriging to improve spatial interpolation of monthly rainfall over complex terrain. *Journal of Hydrometeorogy*, **23**, 561–572.

Norton, C. W., P.-S. Chu, and T. A. Schroeder, 2011: Projecting changes in future heavy rainfall evens for Oahu, Hawaii: A statistical downscaling approach. *Journal of geophysical research*, **116**.

Nullet, D., and M. McGranaghan, 1988: Rainfall enhancement over the Hawaiian islands. *Journal of Climate*, **1**, 847–839.

Rahman, A., and S. Pekkat, 2024: Identifying and ranking of CMIP6-global climate models for projected changes in temperature over Indian subcontinent. *Scientific Reports*, **14 (3076)**.

Rampal, N., and Coauthors, 2024: Enhancing regional climate downscaling through advances in machine learning. *Artificial Intelligence for the Earth Systems*, **3 (2)**, 230 066.

Sanderson, M., 1994: *Prevailing trade winds: weather and climate in Hawai'i*. University of Hawaii Press.

Sanfilippo, K., O. Elison Timm, A. G. Frazier, and T. W. Giambelluca, 2023: Effects of systematic predictor selection for statistical downscaling of rainfall in Hawai‘i. *International Journal of Climatology*, **44**, 571–591.

Schmith, T., 2008: Stationarity of regression relationships: Application to empirical downscaling. *Journal of Climate*, **21 (17)**, 4529–4537.

Virgilio, G. D., and Coauthors, 2022: Selecting CMIP6 GCMs for CORDEX dynamical downscaling model performance, independence, and climate change signals. *Earth's Future*, **10**.