Statistical Downscaling of Climate Models with Deep Learning

Yusuke Hatanaka, Amila Indika, Thomas Giambelluca, Peter Sadowski

March 2024

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, limited historical data exists to train statistical downscaling models for climate, 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 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.

1 Introduction

Climate change is typically modeled using general circulation models (GCMs) that simulate large-scale physical features of atmospheric circulation. Still, the coarse spatial discretization cannot accurately simulate processes at smaller spatial scales, which require the additional step of *downscaling*. Downscaling methods convert coarse-resolution GCM projections to finer-resolution projections [1–5]. 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 Hawaii accurately is a limiting factor [6]. Thus, statistical downscaling is an essential tool for modeling climate change in regions such as Hawai‘i [7–11].

The amount of historical data available for training limits the accuracy of statistical downscaling models [12]. 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 GCM models. They are data-driven and typically fit historical data, with atmospheric data from reanalysis data products and local data from direct measurements, 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 [13]. However, there exist few high-resolution historical data products. Hence, statistical downscaling models typically use simple, linear, statistical models [11]. Some work has been exploring more sophisticated machine learning models [14, 15], but the training data limits the advantage of more complex models. Indeed, machine learning has seen more success in downscaling weather than climate [16] 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; this flexibility enables the modeling of micro-climates but requires abundant historical data at each site. In this work, we propose predicting 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. Hawaii 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 [17, 18], resulting in hyper-local rainfall patterns and micro-climates across the islands. Rainfall is incredibly high on northeast-facing slopes in O'ahu, Maui, and the Big-island, where trade winds force persistent uplift along windward mountain slopes. Second, a high-quality dataset of historical rainfall data is available from the Rainfall Atlas of Hawai'i (RAH) [18], 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.

The rest of the paper is organized as follows. Section 2 describes related work to statistical downscaling approaches for Hawai'i. Section 3 explains the proposed method and those used for experimental comparison. Section 4 summarizes the experimental results, Section 5 discusses the implications and limitations, and Section 6 gives the conclusions. Additional experiments and supplementary information are provided in the Appendix.

2 Related Work

Previous statistical downscaling approaches have relied almost entirely on linear models. Some typical examples are Sanfilippo et al. [11] and Elison Timm et al. [8], which use a combination of linear dimensionality reduction and linear regression to make seasonal rainfall predictions at over 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. Still, the result is that seasonal rainfall at each site is a linear function of the atmospheric features.

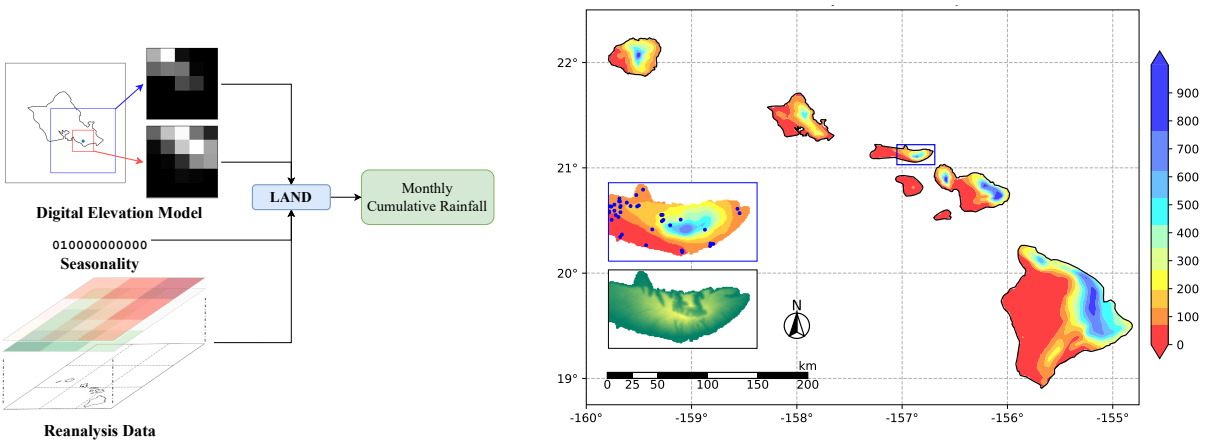
Aside from the modeling limitations of linear models, this approach has other significant limitations. First, the model cannot leverage data from sites with few observations; each site needs enough observations to build a separate site-specific model. Second, because the model is only trained to make 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 bi-linear interpolation or Kriging [19, 20] (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 extrapolates in both the temporal and spatial domain.

3 Methods

This section begins by describing the proposed method, LAND. To be concrete, 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.

3.1 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 — all relevant information about the location is assumed to be represented in the other features, namely the local atmospheric and orographic features. This enables LAND to downscale



(a) Model structure

(b) Predicted monthly rainfall (mm) in January 2000.

Figure 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 differences in the Sun's path. (b) Predicted monthly rainfall (in mm) for January 2000. *Left-Middle*: Zoom of the blue rectangle over the island of Molokaʻi. The blue dots represent the locations of the weather stations in the training data. Weather stations on east Molokaʻi 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 Hawaii.

GCM output to arbitrarily high resolutions, even though the model is trained on sparse observations (See Figure 1b).

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 Oʻahu. Starting from a DEM with 250 m resolution, 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 a location's position relative to the mountain ranges that heavily influence weather in Hawaii. The resolution was chosen to balance the competing goals of including useful features and reducing the tendency to overfit.

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 250,000 parameters. We employ careful hyperparameter tuning and ensemble ten networks with different random initializations to avoid overfitting and reduce the prediction variance. Appendix 8.2 provides the model structure and hyperparameter optimization details.

3.2 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

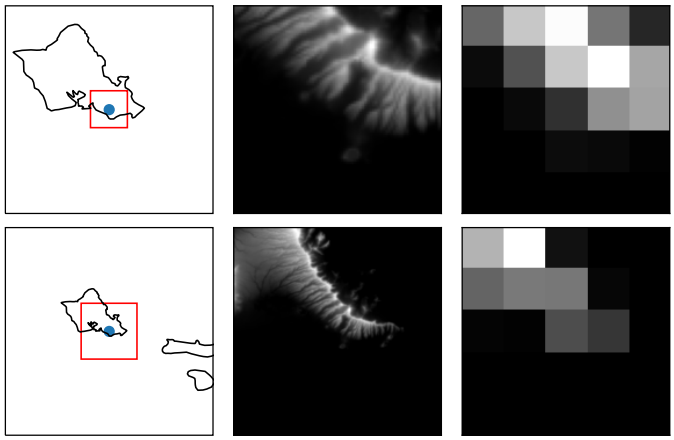


Figure 2: example of local (top) and regional (bottom) orographic features provided to LAND. Top-left: A 20 km region (red square) around a weather station (blue dot) is extracted from the DEM. Top-middle: The region's DEM at full 250 m resolution. Top-right: The region's DEM at coarse resolution; this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

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 8.3), 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 to each weather station site, we refer to this method as site-specific linear regression (SSLR).

To produce maps of predicted rainfall from predictions at individual stations, a GP is used to interpolate between weather 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 8.4), 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 Hawaii [19]. As a result, we obtain a predicted monthly rainfall at any location, achieving the same objective as LAND.

3.3 Datasets

3.3.1 Rainfall Observations

Historical rainfall data was retrieved from the Rainfall Atlas of Hawai'i (RAH) [18], 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 [18] for SSLR but not for LAND, as described below.

3.3.2 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) [21]. The retrieved data consists of monthly mean observations on the 2.5° by 2.5° grid globally since 1948. The complete list of the 16 variables used as the input to the model is available in Appendix 8.1. 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 8.2

3.4 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 used a combination of observational and gap-filled data. This resulted in 355,632 monthly rainfall values from 1,796 weather stations for LAND training and 667,632 observations from 1,102 weather stations for SSLR. We used gap-filled data for training SSLR because otherwise, most sites would not have enough historical data to train a robust site-specific model. The test set consists of all observational data from 2000 to 2012 from any weather station; no gap-filled data was used for evaluation. This results in 56,620 data points from 515 distinct weather stations over 13 years of record.

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 [14]. 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.

3.5 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 observa-

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

4.1 Accuracy

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	

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 8.5.

4.2 Qualitative Comparison

To qualitatively compare the resulting rainfall maps, we plotted the predicted rainfall on a rasterized grid. For example, Figure 3 shows the predicted rainfall maps for six months in 2009, spanning the dry season (May–October) and the rainy season (November–April). While both methods capture the large differences in rainfall between locations, the LAND predictions have fine-grained details absent in the LR+GP predictions. This makes sense, given that the GP kernel uses a smooth RBF kernel to interpolate between stations while LAND uses the local orography. This enables LAND to capture highly localized rainfall patterns that occur against mountain slopes with a precision that is impossible with (LR+GP).

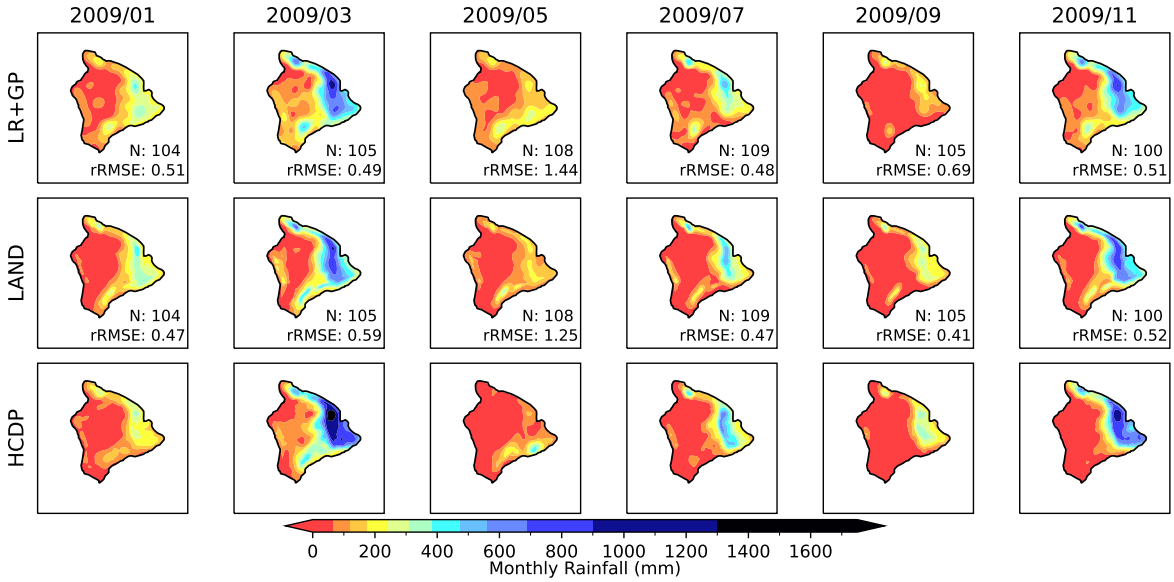


Figure 3: Predicted rainfall maps from LR+GP (top row) and LAND (middle row) for six months in 2009. rRMSE and the number of stations (N) used to calculate the rRMSE are shown for each month. For comparison, the bottom row shows the state-of-the-art rainfall map from the Hawai'i Climate Data Portal (HCDP) [22].

To quantify the advantage of LAND against LR+GP, we focused on 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 rainfall gradient either aligns or opposes the direction of the orography gradient. In other words, it indicates that the rainfall prediction increases or decreases in the same direction as the slope of the land. Though this is not always the case with the actual rainfall pattern, we expect rainfall to be distributed according to the local orography, as orographic lifting is one of the major source of rainfall in Hawai'i. The absolute cosine similarity was calculated at each grid point, and the monthly mean was calculated over all pixels above the sea level. This was repeated over all months in the test period, after which the mean and the standard deviation were calculated, as shown in Table 3.

Table 3: Absolute cosine similarity

Model	**Absolute Cosine Similarity**
LAND	0.731 ± 0.01
LR+GP	0.689 ± 0.01

We found that the absolute cosine similarity is significantly higher for LAND, which indicates alignment between rainfall and the orography in their gradients.

In Figure 4, we provide a close up of March 2009 rainfall predictions and its gradient field on the south of Big Island.

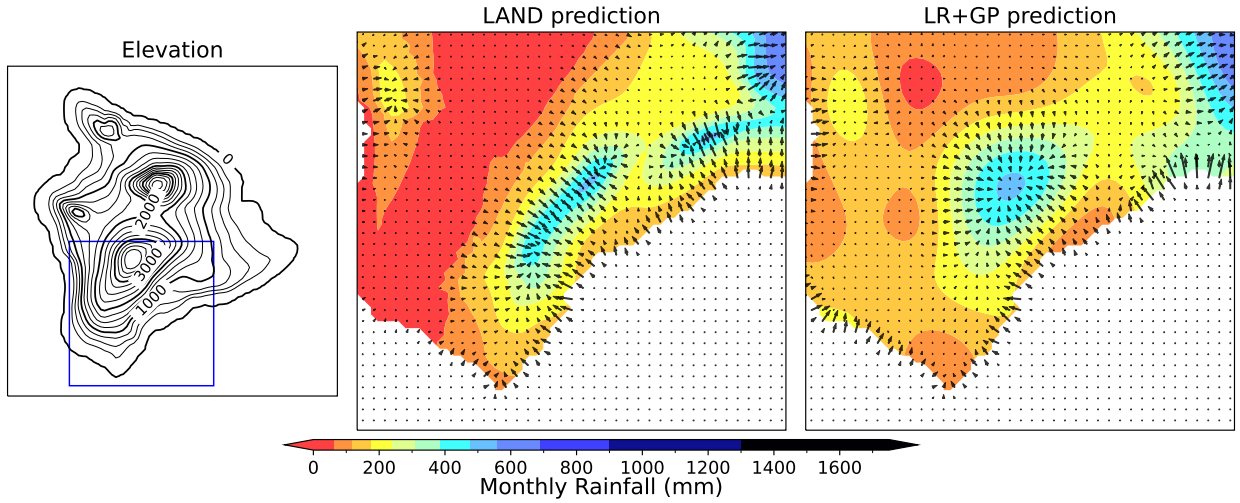


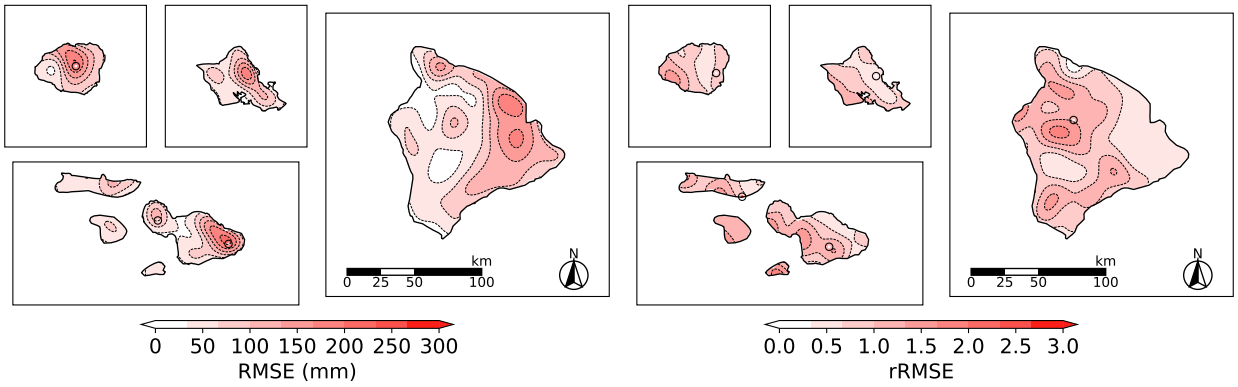
Figure 4: Gradient field of LAND and LR+GP rainfall prediction. LAND predicts rainfall distribution that is consistent with the local orography, whereas LR+GP tends to predict isotropic concentration of rainfall.

We can observe a location where the rainfall gradient of LR+GP is almost isotropic around a center, which does not necessarily match the local orography. This is because GP does not capture the local orographic information. Instead, since LAND directly receives local orography information during training, it has learned a long skinny rainfall distribution along the slope, which is consistent with the local orography.

INFORMAL (revise this paragraph in a a scientific manner. Also requires figures): Another noteworthy observation is the fact that LAND is capable of capturing the Kona max on Big Island. This is atypical to statewide rainfall pattern, as in general, the leeward side is dry due trade wind largely contributing to rainfall on the windward side of the islands, as observed in all but Big Island. In case of Big Island however, massive peaks such as Mauna Loa and Hualalai completely blocks the tradewinds and redirects it such that air flows backward up the slope, causing orographic lift. By accurately capturing such unusual trend, LAND has demonstrated its ability to learn not only statewide rainfall pattern but localized phenomena.

4.3 Spatial Error Distribution

We mapped the spatial distribution of the prediction errors to understand LAND's advantages better. 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. On the other hand, rRMSE tends to be high on the leeward side or inland area on the Big Island, where rainfall is less (and hence the denominator is smaller).

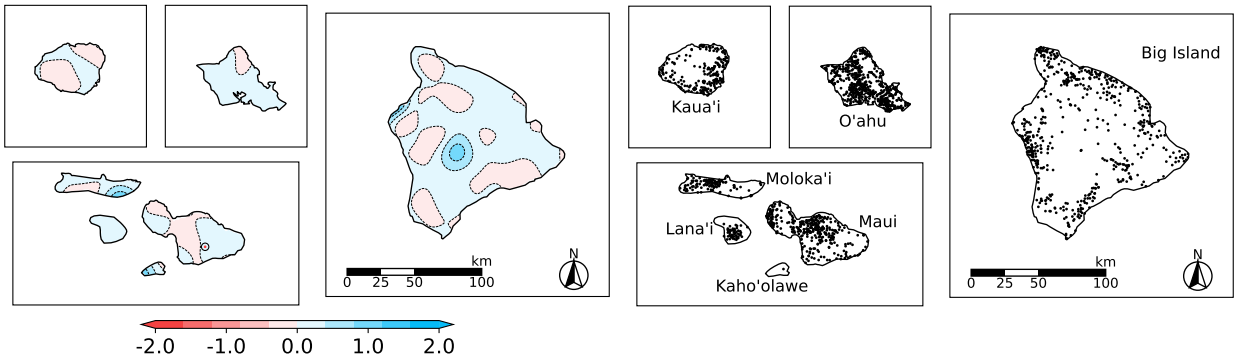


(a) spatial distribution of RMSE

(b) spatial distribution of rRMSE

Figure 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 visualized the distribution 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.



(a) The difference in rRMSE (unitless) at weather stations, interpolated by GP for visualization.

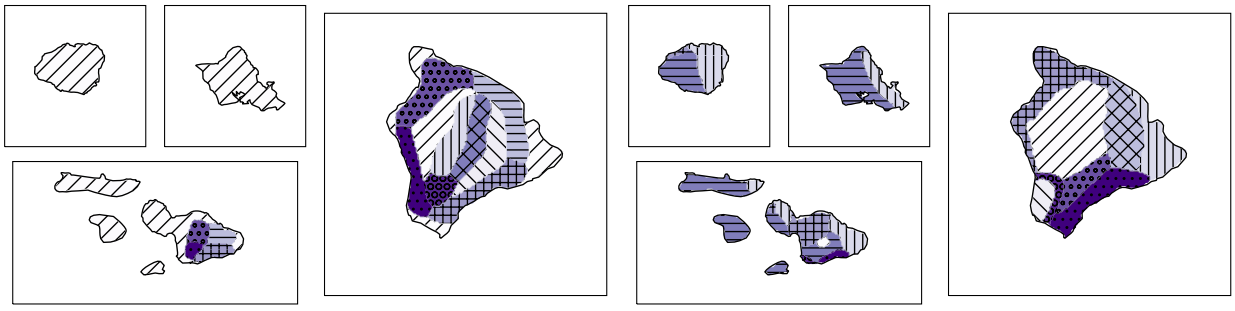
(b) Location of weather stations used for training LAND.

Figure 6: (a) Map showing 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 Molokaʻi, central Big Island, and Kahoʻolawe.

Figures 6a and 6b show that LAND significantly improves in regions with few weather stations. For example, east Molokaʻi, central Big Island, and Kahoʻolawe are regions with few weather stations, and it is in these areas that 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 Kauaʻi, Oʻahu, Molokaʻi, and Kahoʻolawe are clustered together. However, the neural network activation clusters in Figure 7b correspond to known climate regions, for example the windward and leeward sides of Oʻahu. In fact, the NW-facing and

SE-facing regions of multiple islands are 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.



(a) Clustering of each pixel based on raw DEM representation (b) Clustering of each pixel based on neural network activations

Figure 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 configuration of the classes specified by colors and hatches is irrelevant in this context.

4.4 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 4, 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 the variety of climates and geographical characteristics.

Table 4: 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 Leeward		Remote Complex terrain		High rainfall High elevation		High rainfall Windward		High elevation Inland	

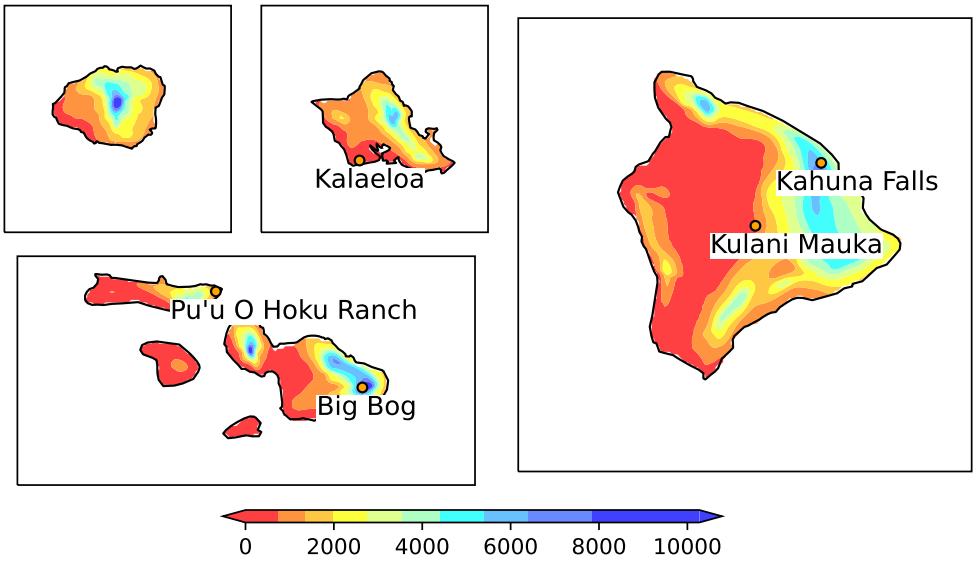


Figure 8: Locations of weather stations for the case study, shown on the map of mean annual rainfall for 1948-1999 (mm), produced by our approach (LAND).

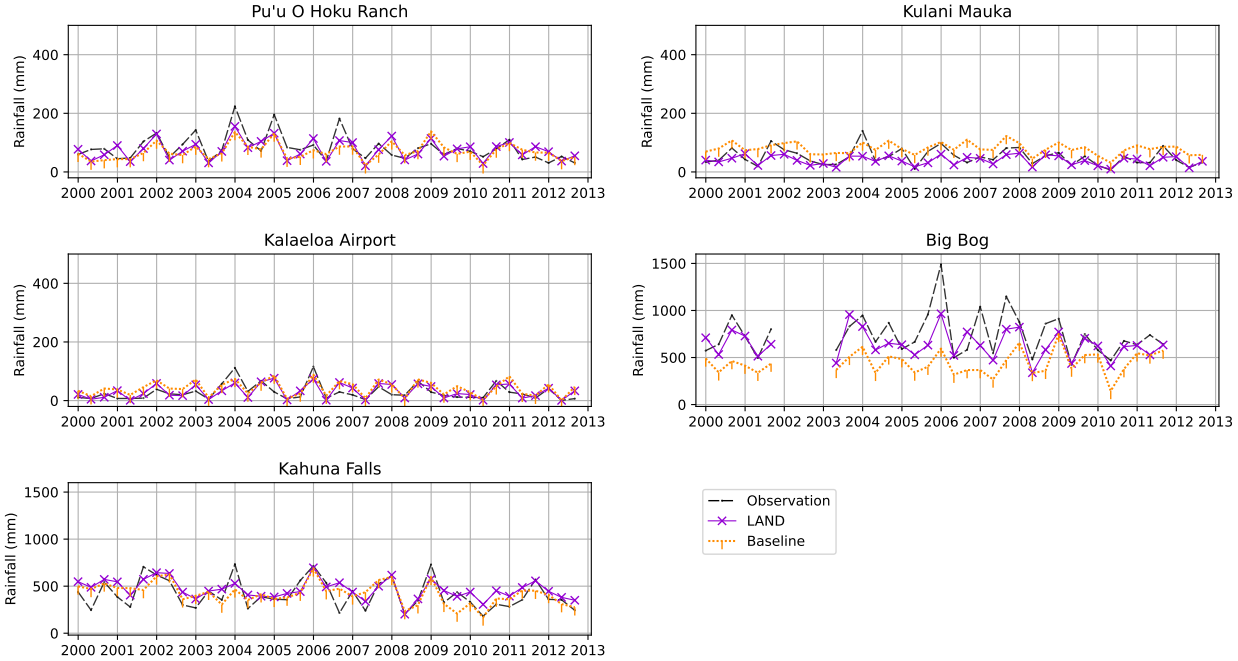


Figure 9: Time series of observed vs. predicted monthly rainfall (in mm) aggregated into three-month bins for clarity.

4.5 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 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.

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.

Table 5: 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	

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 modelled 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 artificial structures/trees near a weather station). 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 4.5 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 [16].

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 Hawaii 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.

7 Acknowledgements

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

References

[1] Moetasim Ashfaq, Deeksha Rastogi, Joy Kitson, Muhammad Adnan Abid, and Shih-Chieh Kao. Evaluation of CMIP6 GCNs over the CONUS for downscaling studies. *Journal of Geophisical Research: Atmospheres*, 127, 2022.

[2] Tolera Abdissa Feyissa, Tamene Adugna Demissie, Fokke Saathoff, and Alemayehu Gebissa. Evaluation of general circulation models CMIP6 performance and future climate change over the omo river basin, ethiopia. *Sustainability*, 15(8), 2023.

[3] Abdul Rahman and Sreeja Pekkat. Identifying and ranking of CMIP6-global climate models for projected changes in temperature over indian subcontinent. *Scientific Reports*, 14(3076), 2024.

[4] Swen Brands. 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, 2022.

[5] Giovanni Di Virgilio, Fei Ji, Eugene Tam, Nidhi Nishant, Jason P. Evans, Chris Thomas, Matthew L. Riley, Kathleen Beyer, Michael R. Grose, Sugata Narsey, and Francois Delage. Selecting CMIP6 GCMs for CORDEX dynamical downscaling model performance, independence, and climate change signals. *Earth's Future*, 10, 2022.

[6] Torben Schmith. Stationarity of regression relationships: Application to empirical downscaling. *Journal of Climate*, 21(17):4529–4537, 2008. URL `https://doi.org/10.1175/2008JCLI1910.1`.

[7] Axel Lauer, Chunxi Zhang, Oliver Elison-Timm, Yuqing Wang, and Kevin Hamilton. Downscaling of climate change in hawaii region using CMIP5 results: On the choice of the forcing fields. *Journal of Climate*, 26, 2013. doi: https://doi.org/10.1175/JCLI-D-13-00126.1.

[8] Oliver Elison Timm, Thomas W. Giambelluca, and Henry F. Diaz. Statistical downscaling of rainfall changes in hawai'i based on the CMIP5 global model projections. *Journal of geophysical research: Atmoshperes*, 120:92–112, 2015.

[9] Oliver Timm and Henry F Diaz. 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, 2009. URL `https://doi.org/10.1175/2009JCLI2833.1`.

[10] O Elison Timm, HF Diaz, TW Giambelluca, and M Takahashi. 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), 2011. URL `https://doi.org/10.1029/2010JD014923`.

[11] Kristen Sanfilippo, Oliver Elison Timm, Abby G. Frazier, and Thomas W. Giambelluca. Effects of systematic predictor selection for statistical downscaling of rainfall in hawai'i. *International Journal of Climatology*, 44:571–591, 2023.

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

[13] Robert Kistler, Eugenia Kalnay, William Collins, Suranjana Saha, Glenn White, John Woollen, Muthuvel Chelliah, Wesley Ebisuzaki, Masao Kanamitsu, Vernon Kousky, Huug van den Dool, Roy Jenne, and Michael Fiorino. The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Sciety*, 82:247–268, 2001.

[14] Neelesh Rampal, Sanaa Hobeichi, Peter B Gibson, Jorge Baño-Medina, Gab Abramowitz, Tom Beucler, Jose González-Abad, William Chapman, Paula Harder, and José Manuel Gutiérrez. Enhancing regional climate downscaling through advances in machine learning. *Artificial Intelligence for the Earth Systems*, 3(2):230066, 2024.

[15] Yusuke M Hatanaka. Machine learning based statistical downscaling for rainfall on Hawaiian islands. Master's thesis, University of Hawai'i at Manoa, 2022.

[16] Yusuke Hatanaka, Yannik Glaser, Geoff Galgon, Giuseppe Torri, and Peter Sadowski. Diffusion models for high-resolution solar forecasts, 2023. URL `https://arxiv.org/abs/2302.00170`.

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

[18] Thomas W. Giambelluca, Qi Chen, Abby G. Frazer, Jonathan P. Price, Yi-Leng Chen, Pao-Shin Chu, Jon K. Eischeid, and Donna M. Delparte. Online Rainfall Atlas of Hawai'i. *Bulltein of the American Meteorological Society*, 94:313–316, 2013.

[19] Matthew P. Lucas, Ryan J. Longman, Thomas W. Giambelluca, Abby G. Frazier, Jared Mclean, Sean B. Cleveland, Yu-Fen Huang, and Jonghyn Lee. Optimizing automated kriging to improve spatial interpolation of monthly rainfall over complex terrain. *Journal of Hydrometeorogy*, 23:561–572, 2022.

[20] Abby G Frazier, Thomas W Giambelluca, Henry F Diaz, and Heidi L Needham. Comparison of geostatistical approaches to spatially interpolate month-year rainfall for the Hawaiian islands. *International Journal of Climatology*, 36(3):1459–1470, 2016.

[21] Eugenia Kalnay, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, John Woollen, et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–472, 1996.

[22] Ryan J. Longman, Matthew P. Lucas, Jared Mclean, Sean B. Cleveland, Keri Kodama, Abby G. Frazier, Katie Kamelamela, Aimee Schriber, Michael Dodge II, Gwen Jacobs, and Thomas W. Giambelluca. The Hawai'i climate data portal. *Bulletin of the American Meteorological Society*, 105:E1074–E1083, 2024.

[23] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019. arXiv:1711.05101.

[24] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts, 2017. arXiv:1608.03983.

[25] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework, 2019. axXiv:1907.10902.

[26] Ryan B. Christianson, Ryan M. Pollyea, and Robert B. Gramacy. Traditional kriging versus modern gaussian processes for large-scale mining data, 2022.

[27] Jacob R Gardner, Geoff Pleiss, David Bindel, Kilian Q Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 2018.

8 Appendix

8.1 Reanalysis Variables

For building downscaling models, we use the following 16 variables listed in Table 6. These variables are consistent with the work by Sanifilippo [11].

Table 6: Climate features used as input to the downscaling models.

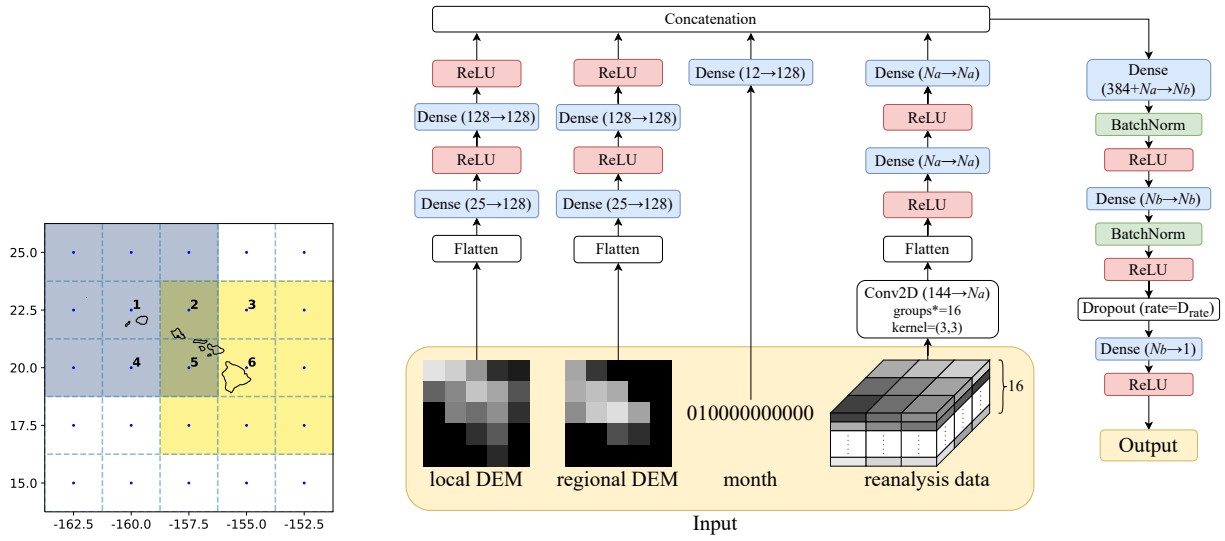
Feature
Geopotential Height at 500hPa
Geopotential Height at 1000hPa
Air temperature difference (1000hPa and 500hPa)
Surface air temperature at 2m
Zonal moisture transport at 700hPa
Zonal moisture transport at 925hPa
Meridional moisture transport 700hPa
Meridional moisture transport 925hPa
Omega
Specific 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

8.2 LAND Details

8.2.1 Model Structure

LAND is a feed-forward neural network that uses local and regional DEMs, month, and reanalysis data as input variables (Figure 10b). As illustrated in Figure 10a, the subset of the composite map corresponding to the closest three-by-three grid cells is extracted and fed to the model at each station. For example, all stations in cell 1 receive the composite map in the blue square, and all stations in cell 6 receive the composite map in the yellow square. This extraction is equally applied to all of the sixteen reanalysis input variables, resulting in data with dimensions $\mathbb{R}^{c \times h \times w}$ where $c = 16$ and $(h, w) = (3, 3)$.

Other possible choices of spatial subsets are $(h, w) \in \{(1, 1), (2, 3), (5, 5)\}$. With $(h, w) = (1, 1)$, each station receives the composite map only on the exact cell to which the station belongs. With $(h, w) = (2, 3)$, all stations receive cells 1 to 6 in Figure 10a, and with $(h, w) = (5, 5)$, all stations receive the entire composite map within Figure 10a. As discussed in a later section, these options were explored as a part of hyperparameter optimization.



(a) (b)

Figure 10: (a) The reanalysis input to the 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. $groups = 16$ is an argument to `torch.nn.Conv2d` and ensures channel-wise independence during convolution.

8.2.2 Data preprocessing and training setup

DEM and rainfall data were divided by 4,000 and 100, respectively, to scale the values so that the variance is approximately one, and reanalysis composite maps were subtracted by the mean and divided by the standard deviation over 1948-2014 for each variable. We used Adam optimizer with decoupled weight decay [23] to minimize the Mean Squared Error (MSE). Cosine annealing with warm restarts [24] was used, where we fixed $T_0 = 10$, T_{mult}=2, and the 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 0.1mm 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. As discussed in the next section, some of the remaining hyperparameters were tuned.

8.2.3 Hyperparameter optimization and model structure search

For hyperparameter optimization, we further split the training dataset into two subsets: data before 1989 (inclusive) for parameter fitting and data after 1989 (exclusive) for validation. 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 7, after which we run an automated hyperparameter search for the Table 8 using Tree-structured Parzen Estimator (TPE) implemented in the `optuna` package for 200 iterations [25]. We then pick the model that resulted in the lowest MSE on the *validation dataset*. Every time we make another choice by hand (from the table Table 7), we repeat the optimization with `optuna` for Table 8. Note that the search is not exhaustive for hand-tuning, meaning not all combinations of choice by hand-tuning were necessarily followed by automated hyperparameter search.

Table 7: 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 [23]}	AdamW
Composite map range 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 8: 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}

8.3 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 only 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 10a). This results in 16 and 96 input variables for the single-cell and grid-cell variants, respectively.

Figure 11 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 [15], 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.

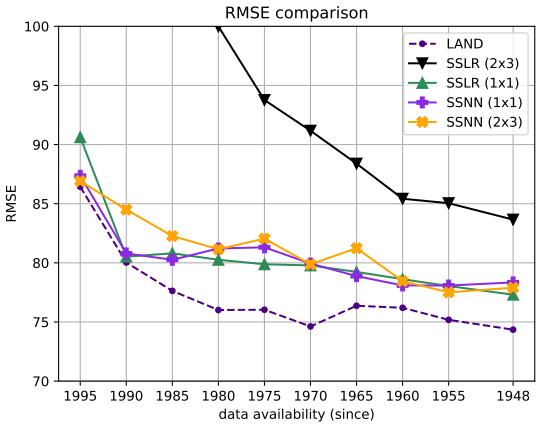


Figure 11: Performance of LAND and site-specific models across different variants.

8.4 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, 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

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

where Θ, length-scale, is another hyperparameter controlling how strongly two points are correlated as a function of distance. As shown above, 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 [26]. 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 [27].

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.

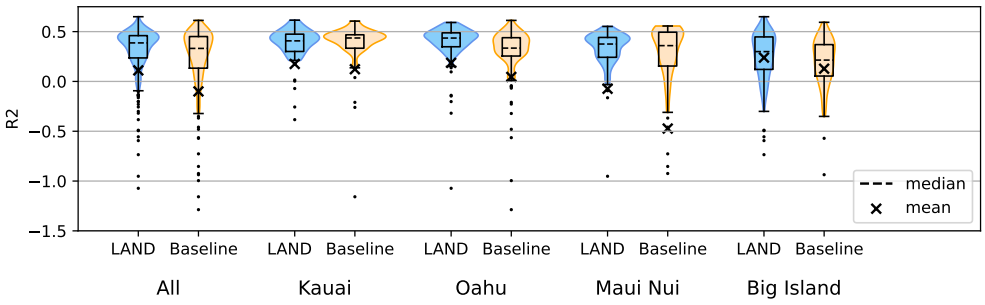
Zero clipping is applied both before and after fitting the Gaussian process model. In other words, predictions from SSLRs are clipped to non-negative values before being fed to the Gaussian process model. Once the Gaussian process interpolates the predictions to a new location, negative values are clipped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

8.5 Error Distribution

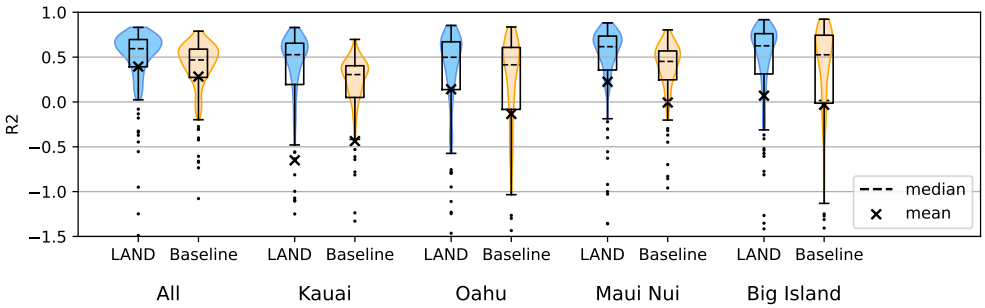
This section discusses the models' performance in terms of R^2. This metric compares the model's performance against a simple mean predictor, in which case, $R^2 = 0$, while $R^2 = 1$ indicates a perfect model with no error. R^2 is robust to absolute measurement unit (as opposed to $RMSE$) or division by small values for

the dry area (as opposed to $rRMSE$). Figure 12a shows the distribution of R^2 per weather station. The violin plots were created by calculating R^2 values at each weather station and then plotting the distribution of R^2 along with the box plot. Black dots represent outliers, 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 12b was created similarly, except the R^2 calculation was done every month.

Regarding Figure 12a, 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 the Figure 12b. 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.



(a) Distribution of R^2 across sites



(b) Distribution of R^2 across months

Figure 12: Distribution of R^2 values on different aggregation criteria.