

DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution

Thomas Vandal
Northeastern University, Civil and
Environmental Engineering
360 Huntington Ave.
Boston, MA 02141
vandal.t@husky.neu.edu

Evan Kodra
risQ Inc.
404 Broadway
Cambridge, MA 02139
evan.kodra@risq.io

Sangram Ganguly
Bay Area Environmental Research
Institute / NASA Ames Research
Center
Moffett Field, CA 94035
sangram.ganguly@nasa.gov

Andrew Michaelis
University Corporation, Monterey
Bay
8 Upper Ragsdale Dr
Monterey Bay, CA 93940
amac@hyperplane.org

Ramakrishna Nemani
NASA Advanced Supercomputing
Division/ NASA Ames Research
Center
Moffett Field, CA 94035
rama.nemani@nasa.gov

Auroop R Ganguly
Northeastern University, Civil and
Environmental Engineering
360 Huntington Ave.
Boston, MA 02141
a.ganguly@neu.edu

ABSTRACT

The impacts of climate change are felt by most critical systems, such as infrastructure, ecological systems, and power-plants. However, contemporary Earth System Models (ESM) are run at spatial resolutions too coarse for assessing effects this localized. Local scale projections can be obtained using statistical downscaling, a technique which uses historical climate observations to learn a low-resolution to high-resolution mapping. Depending on statistical modeling choices, downscaled projections have been shown to vary significantly terms of accuracy and reliability. The spatio-temporal nature of the climate system motivates the adaptation of super-resolution image processing techniques to statistical downscaling. In our work, we present DeepSD, a generalized stacked super resolution convolutional neural network (SRCNN) framework for statistical downscaling of climate variables. DeepSD augments SRCNN with multi-scale input channels to maximize predictability in statistical downscaling. We provide a comparison with Bias Correction Spatial Disaggregation as well as three Automated-Statistical Downscaling approaches in downscaling daily precipitation from 1 degree (100km) to 1/8 degrees (12.5km) over the Continental United States. Furthermore, a framework using the NASA Earth Exchange (NEX) platform is discussed for downscaling more than 20 ESM models with multiple emission scenarios.

CCS CONCEPTS

•Computing methodologies → *Neural networks*; •Applied computing → *Environmental sciences*;

KEYWORDS

Climate Statistical Downscaling, Deep Learning, Daily Precipitation, Super-Resolution

1 INTRODUCTION

Climate change is causing detrimental effects to society's well being as temperatures increase, extreme events become more intense [26], and sea levels rise [25]. Natural resources that society depends on, such as agriculture, freshwater, and coastal systems, are vulnerable to increasing temperatures and more extreme weather events. Similarly transportation systems, energy systems, and urban infrastructure allowing society to function efficiently continue to degrade due to the changing climate. Furthermore, the health and security of human beings, particularly those living in poverty, are vulnerable to extreme weather events with increasing intensity, duration, and frequency [36]. Scientists and stakeholders across areas such as ecology, water, and infrastructures, require access to credible and relevant climate data for risk assessment and adaptation planning.

Earth System Models (ESMs) are physics-based numerical models which run on massive supercomputers to project the Earth's response to changes in atmospheric greenhouse gas (GHG) emissions scenarios. Archived ESM outputs are some of the principal data products used across many disciplines to characterize the likely impacts and uncertainties of climate change [33]. These models encode physics into dynamical systems coupling atmospheric, land, and ocean effects. ESMs provide a large number of climate variables, such as temperature, precipitation, wind, humidity, and pressure, for scientists to study and evaluate impacts. The computationally demanding nature of ESMs limits spatial resolution between 1 and 3 degrees. These resolutions are too coarse to resolve critical physical processes, such as convection which generates heavy rainfall, or to assess the stakeholder-relevant local impacts of significant changes in the attributes of these processes [30].

Downscaling techniques are used to mitigate the low spatial resolution of ESMs through dynamical and statistical modeling. Dynamical downscaling, also referred to as regional climate models (RCMs), account for local physical processes, such as convective and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD'17, August 13–17, 2017, Halifax, NS, Canada.

© 2017 ACM. ISBN 978-1-4503-4887-4/17/08...\$15.00

DOI: <http://dx.doi.org/10.1145/3097983.3098004>

vegetation schemes, with sub-grid parameters within ESM boundary conditions for high-resolution projections. Like ESMs, RCMs are computationally demanding and are not transferable across regions. In contrast, the statistical downscaling (SD) technique learns a functional form to map ESMs to high resolution projections by incorporating observational data. A variety of statistical and machine learning models, including linear models [16], neural networks [6], and support vector machines [13], have been applied to SD and is discussed further in section 4 (Related Work). Despite the availability of many techniques, we are not aware of any SD method which explicitly captures spatial dependencies in both low-resolution and high-resolution climate data. Furthermore, traditional methods require observational data at the high-resolution target, meaning that regions with little observation data, often the poorest regions which are most effected by climate change, are unable to receive downscaled climate data needed for adaptation.

The lack of explicit spatial models in SD of ESMs motivated us to study the applicability of computer vision approaches, most often applied to images, to this problem. More specifically, advances in single image super-resolution (SR) correspond well to SD, which learns a mapping between low- and high-resolution images. Moreover, as SR methods attempt to generalize across images, we aim to provide downscaled climate projections to areas without high-resolution observations through what may be thought of as transfer learning. Though we will discuss this topic further in section 4 (Related Work), we found that super-resolution convolutional neural networks were able to capture spatial information in climate data to improve beyond existing methods.

Lastly, we present a framework using our super-resolution approach to downscale ensemble ESMs over the Continental United States (CONUS) at a daily temporal scale for four emission scenarios by using NASA's Earth Exchange (NEX) platform. NEX provides a platform for scientific collaboration, knowledge sharing and research for the Earth science community. As part of NEX, along with many other earth science data products, NASA scientists have already made monthly downscaled ESMs for CONUS up to the year 2100 at 30 arc seconds (NEX-DCP30) that are openly available to the public. However, the downscaling methodology, bias correction spatial disaggregation, has limitations and the monthly scale reduces the applicability to studying extreme events. The improvement of such data products is vital for scientists to study local impacts of climate change to resources society depends on.

1.1 Key Contributions

The key contributions are as follows:

- We present DeepSD, an augmented stacked super-resolution convolutional neural network for statistical downscaling of climate and earth system model simulations based on observational and topographical data.
- DeepSD outperforms a state-of-the-art statistical downscaling method used by the climate and earth science communities as well as a suite of off-the-shelf data mining and machine learning methods, in terms of both predictive performance and scalability.
- The ability of DeepSD to outperform and generalize beyond grid-by-grid predictions suggests the ability to leverage cross-grid information content in terms of similarity of

learning patterns in space, while the ability to model extremes points to the possibility of improved ability beyond matching of quantiles. Taken together, this leads to the new hypothesis that methods may be able to use spatial neighborhood information to predict in regions where data may be sparse or low in quality.

- For the first time, DeepSD presents an ability to generate, in a scalable manner, downscaled products from model ensembles, specifically, simulations from different climate modeling groups across the world run with different emissions trajectories and initial conditions.
- DeepSD provides NASA Earth Exchange (NEX) a method of choice to process massive climate and earth system model ensembles to generate downscaled products at high resolutions which can then be disseminated to researchers and stakeholders.

1.2 Organization of the Paper

The remainder of the paper is organized as follows. Section 2 (Earth Science Data) presents necessary data used for SD along with their associated challenges. Section 3 (Statistical Downscaling) discusses the problem of SD. Section 4 (Related Work) discusses techniques previously applied to SD along with an overview of super-resolution methods. Furthermore, we discuss the relationships between images and climate data. Section 5 (Methodology) presents DeepSD, the augmented stacked super-resolution convolutional neural network formulation. In section 6 (Experiments) we compare our method to another SD technique and three off-the-shelf machine learning approaches and outline the process by which we will scale our method to many climate model simulations. In section 7 (conclusion) we briefly discuss results, limitations, and future work.

2 EARTH SCIENCE DATA

Earth science data stems from a variety of areas, including climate simulations, remote sensing through satellite observations, and station observations. The spatio-temporal nature of such data causes heavy computational and storage challenges. For instance, a single climate variable at the daily temporal and 4km spatial scales over only the United States requires 1.2GB of storage. Multiplying this effect over a large number of variables, including precipitation, temperature, and wind, globally creates high storage and processing requirements. Furthermore, analysis of these complex datasets require both technical and domain expertise.

ESM outputs, as discussed previously, are one form of earth science data which is crucial to the understanding of our changing climate. The most recent ESMs are a product of the fifth phase of the Coupled Model Intercomparison Project which simulate the climate through a dynamical system coupling effects from the atmosphere, land, and ocean [33]. However, it is well understood that holes in these models exist, including low-resolution and lack of model agreement, particularly for precipitation [30].

We can harness information in observational datasets in order to learn statistical models mapping ESM outputs to a higher resolutions. Observational datasets are available through a variety of sources, including satellite observations, station observations, and

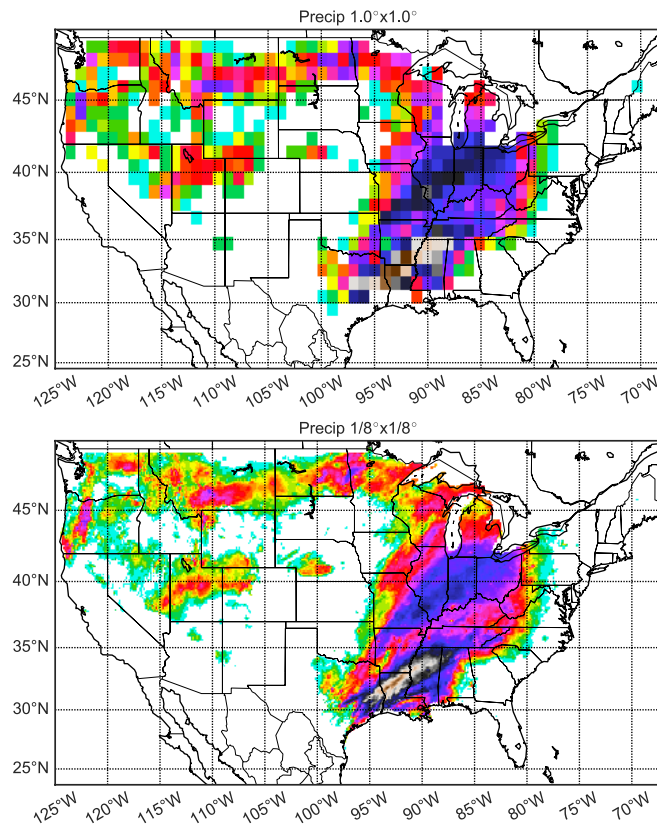


Figure 1: Prism Observed Precipitation: A) Low resolution at 1.0° ($\sim 100\text{km}$). B) High resolution at $1/8^\circ$ ($\sim 12.5\text{km}$).

a mixture of both, namely reanalysis datasets. Often, SD models are built to downscale ESMs directly to an observational station while others aim to downscale to a grid based dataset. Gridded observational datasets are often built by aggregating station observations to a defined grid. For example, in our application, we obtain precipitation through the PRISM dataset at a 4km daily spatial resolution which aggregates station observations to a grid with physical and topographical information [8]. We then upscale the precipitation data to $1/8^\circ$ (12.5 km) as our high-resolution observations. Following, we upscale further to 1° corresponding to a low-resolution precipitation, as applied in [28]. The goal is then to learn a mapping between our low-resolution and high-resolution datasets.

Furthermore, topography has large effects on weather and climate patterns with lower temperatures, more precipitation, and higher winds [8, 9]. Taking advantage of the valuable topographical information at different scales, where $1/2^\circ$ may capture large scale weather patterns while $1/8^\circ$ spatial resolution can capture high-resolution precipitation biases.

Each of the earth science data products discussed inherently possess rich spatial dependencies, much like images. However, traditionally statistical downscaling methods, particularly regression based models, vectorize spatial data, removing this spatial structure. While colored images contain channels consisting of, for example, red, green, and blue, climate data may be represented analogously such that the channels correspond to climate variables

and topographical data. Similar approaches have been applied to satellite datasets for image classification [2] and resolution enhancement [44]. Though climate data is more complex than images due to its dynamics and chaotic nature, we propose that this representation allows scientists to approach the data in an unconventional manner and apply augmented models developed for image processing.

3 STATISTICAL DOWNSCALING

SD is the problem of mapping a low-resolution climate variable to a high-resolution projection. This mapping, which must transform a single grid point to multiple points is an ill-posed problem, one with many possible solutions (see Figure 1). However, we can mitigate the ill-posed problem by including static high-resolution topography data in conjunction with other low-resolution climate variables. We learn the SD model using observed climate datasets and then infer downscaled ESM projections. Spatial and temporal non-stationarity of the changing climate system challenges traditional statistical techniques. Downscaling precipitation further challenges these methods with sparse occurrences and skewed distributions. The combination of an ill-posed problem, uncertainty in the climate system, and data sparsity propagates uncertainty in downscaled climate projections further.

4 RELATED WORK

As mentioned previously, SD has a rich and expansive history in the climate community. SD consists of three fundamental categories: regression models and weather classification schemes which improve spatial resolution while weather generators increase temporal resolution (ie. monthly to daily) [40]. As our interest is in increasing spatial resolution we will review regression methods and weather classification.

Regression methods applied to SD are wide in scope, both linear and non-linear, and vary based on the specific climate variable and temporal scale. For instance, downscaling daily precipitation, which we will focus on, relies on a sparse observational dataset where few days contain rainfall while the amount of rainfall in those days follow a skewed distribution. Automated Statistical Downscaling (ASD) presents a traditional framework for this problem where a classification model is first used to classify days with precipitation followed by a regression to estimate the amount [16]. Similar approaches, among others, include quantile regression neural networks [6], bayesian model averaging[45], and expanded downscaling[3]. Each of these regression models learns a statistical relationship between observed low- and high-resolution pairs and is then applied to ESMs. Another widely used approach is Bias Corrected Spatial Disaggregation (BCSD), which begins by bias correcting an ESM to match the distribution of the high-resolution observed dataset followed by interpolation and spatial scaling to correct for local biases[34, 41]. Though BCSD is a simple approach, it has been shown to perform well compared to more complex methods [4, 23]. Furthermore, we have shown that BCSD performs similarly, or better, when compared to off-the-shelf ASD approaches [37].

Weather classification methods take a different approach to statistical downscaling through nearest neighbor estimates, grouping

weather events into similar types. Given a set of observed low- and high-resolution pairs, one can compute a distance measure between an ESM and the low-resolution observations to select the nearest high-resolution estimate. Constructed analogues furthers the method by performing a regression on a group of the nearest neighbor estimates[17]. More advances, but similar approaches, have recently been presented, including Hierarchical Bayesian inference models[22].

While the approaches discussed above are often sufficient in downscaling means, they tend to fail at downscaling extreme events. For instance, ASD approaches perform reasonably well at downscaling average precipitation [16] but performs poorly at the extremes [4]. As discussed by Bürger et al. [4] and Mannshardt-Shamseldin et al. [21], as well as others, specific approaches to downscaling extremes are often required. These specialized approaches, such as those using Generalized Extreme Value theory, have been developed for this purpose [15, 21]. Ideally, a single approach to downscaling leveraging all available information would capture both averages and extremes, giving the user a more credible dataset.

To our knowledge little work has been attempted to explicitly capture spatial properties for improving downscaled projections. As computer vision approaches are built to exploit the spatial structure of images, we are motivated to understand the applicability of such methods to climate datasets. As introduced previously, we represent climate variables as channels, analogous to images, and model them similarly. However, we note that this presents an analogy and not a direct correspondence.

Using the analogy between climate datasets and images, we can relate statistical downscaling to image super-resolution, where one aims to learn a mapping from low- to high-resolution image pairs. Specifically, single image super-resolution (SR), as the name suggests, increases the resolution of a single image, rather than multiple images, from a scene.

The most successful approaches to SR have been shown to be patch based (or example-based) techniques, achieving state-of-the-art performance [10, 35, 39]. Originally proposed by Glasner et al. [14], patch based methods exploit self-similarity between images to produce exemplar patches. This approach has evolved into different variations of nearest neighbor techniques between low- and high-resolution patches through what is known as dictionary learning [7, 11, 43]. Dictionary learning approaches to SR are analogous to those presented by weather classification SD schemes. Furthermore, approaches including kernel regression [42], random forests[31], and anchored neighborhood regression [35], have been proposed for SR to improve accuracy and speed, all related to methods presented in SD literature [5, 13, 32]. Sparse-coding techniques, a form of dictionary learning, have recently shown state-of-the-art results in both speed and accuracy [35].

Convolutional neural networks were recently presented as a generalization of sparse coding, improving upon past state-of-the-art performances[10, 39]. The sparse coding generalization, non-linearity, network flexibility, and scalability to large datasets presents an opportunity to apply Super Resolution Convolutional Neural Networks to SD[10].

5 METHODOLOGY

This section begins by describing and formulating Super-Resolution Convolutional Neural Networks (SRCNN), as presented by [10]. We then introduce a stacked SRCNN architecture such that the output of one SRCNN is the input to the following SRCNN. DeepSD, the adaptation of a stacked SRCNNs to SD, is then introduced.

5.1 Super-resolution CNN

SR methods, given a low-resolution (LR) image, aim to accurately estimate a high-resolution image (HR). As presented by Dong et al. [10], a CNN architecture can be designed to learn a functional mapping between LR and HR using three operations, patch extraction, non-linear mappings, and reconstruction. The LR input is denoted as X while the HR label is denoted as Y .

A three layer CNN is then constructed as follows to produce a high resolution estimate and presented in Figure 2. Layer 1 is formulated as

$$F_1(X) = \max(0, W_1 * X + B_1),$$

where ‘ $*$ ’ is the convolution operation and the $\max$ operation applies a Rectified Linear Unit [24] while W_1 and B_1 are the filters and biases, respectively. W_1 consists of n_1 filters of size $c \times f_1 \times f_1$. The filter size, $f_1 \times f_1$, operates as an overlapping patch extraction layer where each patch is represented as a high-dimensional vector.

Correspondingly, layer 2 is a non-linear operation such that

$$F_2(X) = \max(0, W_2 * F_1(X) + B_2)$$

where W_2 consists of n_2 filters of size $n_1 \times f_2 \times f_2$ and B_2 is a bias vector. This non-linear operation maps high-dimensional patch-wise vectors to another high-dimensional vector.

A third convolution layer is used to reconstruct a HR estimate such that

$$F(X) = W_3 * F_2(X) + B_3.$$

Here, W_3 contains 1 filter of size $n_2 \times f_3 \times f_3$. The reconstructed image $F(X)$ is expected to be similar to the HR image, Y

This end-to-end mapping then requires us to learn the parameters $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$. A Euclidean loss function with inputs $\{X_i\}$ and labels $\{Y_i\}$ is used where the optimization objective is defined as:

$$\underset{\Theta}{\operatorname{argmin}} \sum_{i=1}^n \|F(X_i; \Theta) - Y_i\|_2^2 \quad (1)$$

such that n is the number of training samples (batch size).

We note that convolutions in layers 1, 2, and 3 decrease the image size depending on the chosen filter sizes, f_1, f_2 , and f_3 . At test time, padding using the replication method is applied before the convolution operation to ensure the size of the prediction and ground truth correspond. During training, labels are cropped such that Y and $F(X_i; \Theta)$, without padding, are of equal size.

5.2 Stacked SRCNN

Traditional SR methods are built for resolution enhancements of factors from 2 to 4 while statistical downscaling conservatively requires resolution increases of factors from 8 to 12. Rather than enhancing resolution directly to 8-12x, as SR applications typically do, we take an alternative approach. To achieve such a large resolution improvement, we present stacked SRCNNs such that each

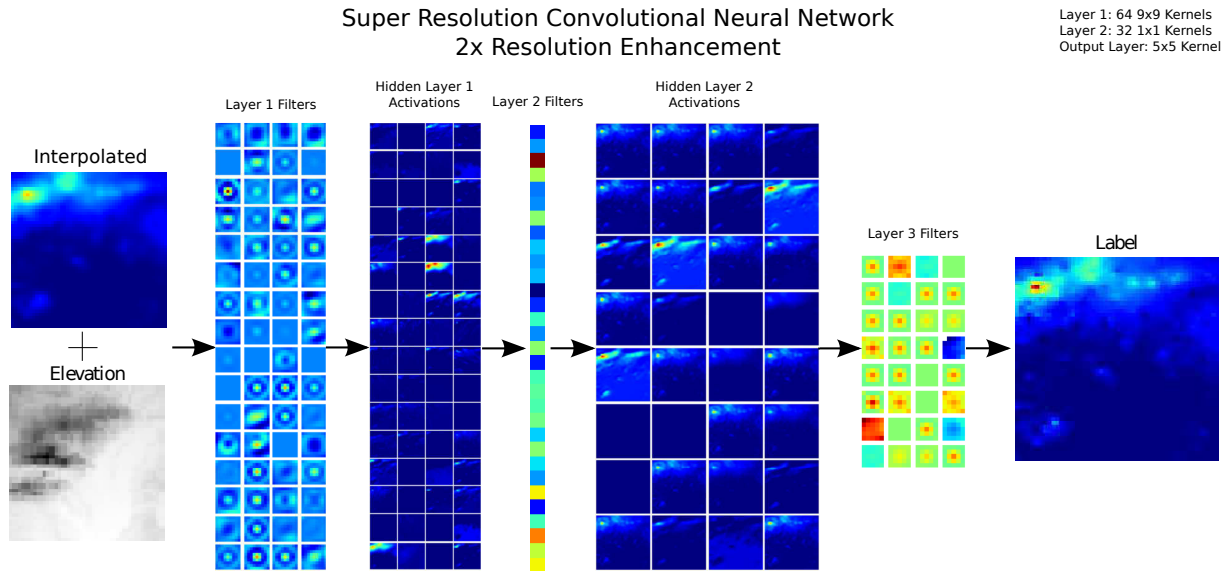


Figure 2: Augmented SRCNN Architecture. From the left to right: Precipitation and Elevation sub-image pair, filters learned in layer 1, layer 1 activations, layer 2 filters, layer 2 activations, layer 3 filters, and HR precipitation label.

SRCNN increases the resolution by a factor of s . This approach allows the model to learn spatial patterns at multiple scales, requiring less complexity in the spatial representations. The approach of stacking networks has been widely used in deep learning architectures, including stacked denoising autoencoders [38] and stacked RBMs for deep belief networks [18]. However, contrary to the above networks where stacking is applied in an unsupervised manner, each SRCNN is trained independently using their respective input/output resolutions and stacked at test time.

A similar approach using cascading super-resolution networks showed positive results for upscaling factors below 4 [39], however through experimentation we found that cascading SRCNNs performed worse than stacked SRCNNs. The ability of arbitrarily upscaling ground truth images to lower resolution allows for input/output pairs to be produced at multiple scales to train stacked SRCNNs. However, while training a cascading model, the output of each SRCNN is the input to the following SRCNN, which may be leading to unnecessary error propagation through the network.

5.3 DeepSD

We now present DeepSD, an augmented and specific architecture of stacked SRCNNs, as a novel SD technique. When applying SR to images we generally only have a LR image to estimate a HR image. However, during SD, we may have underlying high-resolution data coinciding with this LR image to estimate the HR images. For instance, when downscaling precipitation we have two types on inputs including LR precipitation and static topographical features such as HR elevation and land/water masks to estimate HR precipitation. As topographical features are known beforehand at very high resolutions and generally do not change over the period of interest they can be leveraged at each scaling factor. As done when training stacked SRCNNs, each SRCNN is trained independently with its associated input/output pairs. As presented in figure 3,

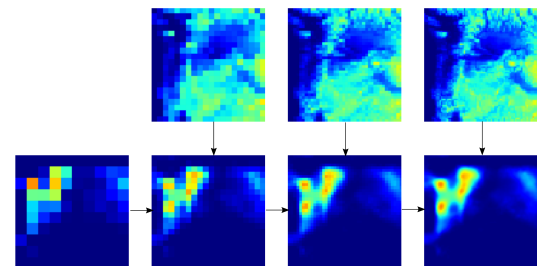


Figure 3: Layer by layer resolution enhancement from DeepSD using stacked SRCNNs. Top Row: Elevation, Bottom Row: Precipitation. Columns: 1.0° , $1/2^\circ$, $1/4^\circ$ and $1/8^\circ$ spatial resolutions.

inference is executed by starting with the lowest resolution image with its associated HR elevation to predict the first resolution enhancement. The next resolution enhancement is estimated from the previous layer's estimate and its associated HR elevation. This process is repeated for each trained SRCNN. Figure 3 illustrates this process with a precipitation event and its various resolution improvements. We see that this stacked process allows the model to capture both regional and local patterns.

6 APPLICATION OF DEEPSD

Though high resolution precipitation is crucial to climate adaptation, it makes up two of the four major holes in climate science [30]. Furthermore, both statistical and dynamical downscaling approaches have been shown to add little information beyond coarse ESMs when applied to precipitation [4, 19]. This motivates our application to downscale daily precipitation over the CONUS, a region where data is credible and abundant at high resolutions.

As presented above, we use daily precipitation from the PRISM dataset [8] and elevation from Global 30 Arc-Second Elevation Data Set (GTOPO30) provided by the USGS. These datasets are used to train and test DeepSD, which we compare to BCSD, a widely used statistical downscaling technique, as well as three off-the-shelf machine learning regression approaches. The years 1980 to 2005 were used for training (9496 days) while the years 2006 and 2014 (3287 days) were used for testing. Lastly, we present a scalable framework on the NASA Earth Exchange (NEX) platform to downscale 20 GCMs for multiple emission scenarios.

6.1 Training DeepSD

Our experiments downscale daily precipitation from 1.0° to $1/8^\circ$, an 8x resolution enhancement, using three SRCNN networks each providing a 2x resolution increase ($1.0^\circ \rightarrow 1/2^\circ \rightarrow 1/4^\circ \rightarrow 1/8^\circ$). For comparison and experimentation purposes, we also trained a single network with 8x resolution increase which we denote as SRCNN in the results.

Data preprocessing. Data for a single day at the highest resolution, $1/8^\circ$, covering CONUS is an “image” of size 208x464. Precipitation and elevation are used as input channels while precipitation is the sole output. Images are obtained at each resolution through up-sampling using a bicubic interpolation. For instance up-sampling to 1.0° decreases the image size from 208x464 to 26x58. Precipitation features for the first SRCNN, downscaling from 1.0° to $1/2^\circ$, are first up-sampled to 1.0° and then interpolated for a second time to $1/2^\circ$ in order to correspond to the output size of 52x116. This process is subsequently applied to each SRCNN depending on its corresponding resolution. During the training phase, 51x51 sub-images are extracted at a stride of 20 to provide heterogeneity in the training set. The number of sub-images per year (1095, 9125, and 45,625) increase with resolution. Features and labels are normalized to zero mean and unit variance. Precipitation values are only available over land so we set each null value to a sufficiently low value of -5 which is then masked after downscaling accordingly.

Training Parameters. All SRCNNs are trained with the same set of parameters, selected using those found to work well by Dong et al. [10]. Layer 1 consists of 64 filters of 9x9 kernels, layer 2 consists of 32 filters of 1x1 filters, and the output layer uses a 5x5 kernel (see Figure 2). Higher resolution models which have a greater number of sub-images may gain from larger kernel sizes and an increased number of filters. Each network is trained using Adam optimization [20] with a learning rate of 10^{-4} for the first two layers and 10^{-5} for the last layers. Each model was trained for 10^7 iterations with a batch size of 200. Tensorflow [1] was utilized to build and train DeepSD. Training harnessed three Titan X GPUs on an NVIDIA DIGITS DevBox by independently training each SRCNN. Inference was then executed sequentially on a single Titan X GPU on the same machine.

6.2 Comparison

6.2.1 State-of-the-Art Methods. Bias Correction Spatial Disaggregation (BCSD) [41] is a simple but effective method for statistical downscaling. Many studies have compared BCSD to a variety of

Model	Bias (mm/day)	Corr	RMSE (mm/day)	Skill	Runtime (secs)
Lasso	0.053	0.892	2.653	0.925	1297
ANN	0.049	0.862	3.002	0.907	2015
SVM	-1.489	0.886	3.205	0.342	27800
BCSD	-0.037	0.849	4.414	0.955	13
SRCNN	-0.699	0.894	2.949	0.833	24
DeepSD	0.022	0.914	2.529	0.947	71

Table 1: Comparison of predictive ability between all six methods for 1000 randomly selected locations in CONUS. Runtime is computed as the amount of time to downscale 1 year of CONUS.

other downscaling methods and have found good results in estimating the underlying distribution of precipitation [4]. In our experiments we apply the daily BCSD technique [34] to precipitation over CONUS. First, the high-resolution precipitation is linearly interpolated to the low-resolution grid. Contrary to applying BCSD to ESMs, we are not required to perform quantile mapping as the distributions are identical. Next, the low-resolution precipitation is interpolated back to high-resolution such that the fine-grained information is lost. Then, scaling factors are computed by dividing high-resolution observations with the interpolated data over the training set (1980-2005). Lastly, the interpolated data is multiplied by the scaling factors to provide downscaled projections. For a more detailed description of this implementation of BCSD see [34]. The projections over the test set (2006-2014) are then used for comparison to BCSD.

A second set of methods, Automated-Statistical Downscaling (ASD) [16], is applied to compare a variety of regression techniques to DeepSD. ASD consists of two steps for downscaling precipitation: 1. Classifying rainy/non-rainy days (mm), 2. Estimating total precipitation on rainy days. Hence, this approach requires both classification and regression methods. We compare three ASD approaches using logistic and lasso regression, support vector machine (SVM) classifier and regression, and artificial neural network (ANN) classifier and regression. The Lasso penalty parameter at each location was chosen using 3-fold cross-validation. The SVMs were trained with a linear kernel and a penalty parameter of 1.0. Each ANN consists of a single layer of 100 units connected with a sigmoid function. A 9 by 9 box for the LR precipitation surrounding the downscaled location is selected as features. Each downscaled location requires individually optimized parameters making the process computationally intensive and complex. Hence, we randomly selected 1000 locations to downscale as a trade-off between complexity and statistical certainty around our results. All features and labels are normalized to zero mean and unit variance.

6.2.2 Daily Predictability. DeepSD’s ability to provide credible projections is crucial to all stakeholders. While there are many facets to statistical downscaling, we use a few key metrics to show DeepSD’s applicability. Root mean square error (RMSE) and Pearson’s correlation are used to capture the predictive capabilities of the methods. Figure 4 maps this RMSE (mm/day) for each location. Bias, the average error, presents the ability to estimate the mean while a skill score metric, as presented in [27], is used to measure

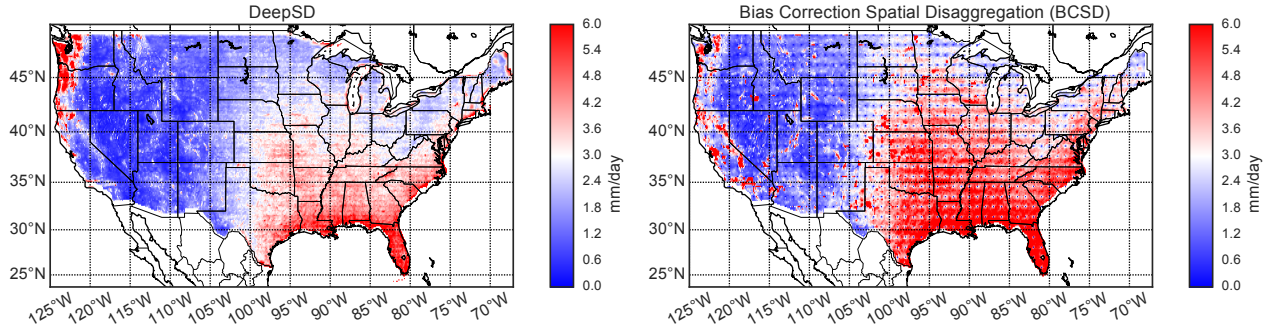


Figure 4: Daily Root Mean Square Error (RMSE) computed at each location for years 2006 to 2014 (test set) in CONUS for Left) DeepSD and Right) BCSD. Red corresponds to high RMSE while blue corresponds to low RMSE.

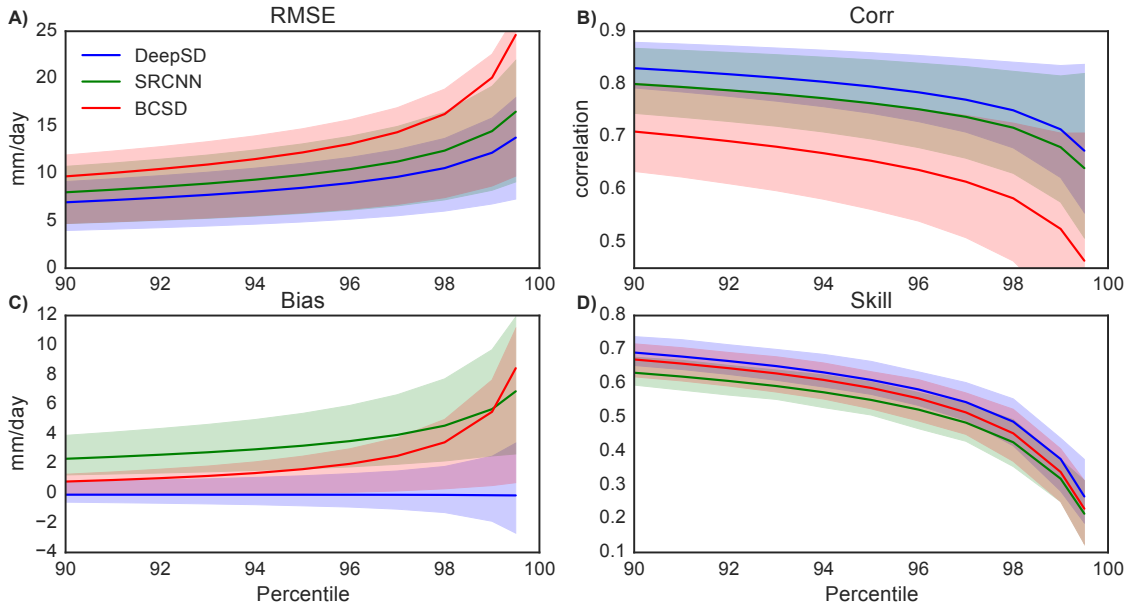


Figure 5: Comparison of DeepSD, SRCNN, and BCSD for increasingly extreme precipitation. At each location in CONUS, all precipitation events above a percentile threshold (x-axis) are selected. Percentile thresholds between 90 and 99.9 are used. A) RMSE, B) Correlation, C) Bias, and D) Skill are computed at each location and averages over CONUS. The confidence bounds of each metric are taken from the 25th and 75th quantiles.

distribution similarity. Skill is computed as

$$skill = \sum_{i=1}^n \text{minimum}(Z_o^{(i)}, Z_m^{(i)}) \quad (2)$$

such that Z_o and Z_m are the observed and DeepSD's empirical probability density function while n is the number of bins. Hence, the skill score is between 0 and 1 where 1 is the best.

Our first experiment compares six approaches, DeepSD, SRCNN (DeepSD w/o stacking), BCSD, Lasso, SVM, and ANN, on their ability to capture daily predictability, presented in Table 1. The four metrics discussed above are computed and averaged over the 1000 randomly selected locations in CONUS where ASD methods were trained. We find that DeepSD outperforms the other approaches in terms of bias, correlation, and RMSE and closely behind BCSD in

Season	Model	Bias (mm/day)	Corr	RMSE (mm/day)	Skill ¹
DJF	BCSD	0.02	0.89	2.36	0.95
	DeepSD	-0.03	0.95	1.53	0.94
JJA	BCSD	0.01	0.78	4.15	0.92
	DeepSD	-0.05	0.86	3.29	0.91
MAM	BCSD	0.01	0.87	3.02	0.94
	DeepSD	-0.03	0.93	2.29	0.93
SON	BCSD	0.01	0.87	3.27	0.94
	DeepSD	-0.04	0.93	2.31	0.94

Table 2: Comparison of Predictive Ability between DeepSD and BCSD for each season, Winter, Summer, Spring, and Fall. Values are computed at each location in CONUS and averaged.

terms of skill. We also find that the stacking performed by DeepSD provides a large performance improvement beyond a single SRCNN network with an 8x resolution increase. Furthermore, we find that SVM performs poorly in testing while having the longest runtime. Similarly, the least complex ASD method, Lasso, outperforms the more complex ANN. As expected, BCSD, a method built around estimating the underlying distribution, does well in minimizing bias and estimating the underlying distribution. For these reasons, in conjunction with our previous findings [37], the remaining experiments will limit the methods to DeepSD and BCSD.

In the next experiment compare DeepSD and BCSD, the two scalable and top performing methods from the previous experiment, with each metric over CONUS. Each metric is computed per location and season using the daily observations and downscaled estimates then averaged over CONUS, as presented in Table 2. We find that DeepSD has high predictive capabilities for all seasons, higher correlation and lower RMSE, when compared to BCSD. Similar results are shown in Figure 4 where DeepSD has a lower RMSE than BCSD for 79% of CONUS. Furthermore, we find each method’s ability to estimate the underlying distribution well with low bias, < 0.5 mm/day, and a high skill score of ~ 0.98. As BCSD is built specifically to minimize bias and match the underlying distribution, DeepSD’s performance is strong. Overall, DeepSD outperforms BCSD for the chosen set metrics.

6.2.3 Predicting Extremes. As discussed in section 4, downscaling both averages and extreme events with a single method is challenging. Our last experiment tests this challenge by comparing DeepSD’s ability to estimate extreme precipitation events when compared to BCSD, an approach shown to perform well [4]. A varying quantile threshold approach is used to test each methods ability to capture extreme events. For instance, given a downscaled location we compute RMSE, correlation, bias, and skill for all precipitation events greater the 90th percentile. This is done for a range of percentiles between 90 and 99.9 and averaged over all locations in CONUS. Along with the mean, we select the 25th and 75th quantiles of each metric over CONUS and plot them as confidence bounds in Figure 5. Figure 5 presents BCSD’s loss of predictive capability when compared to DeepSD. We find that BCSD over-estimates extremes at upper quantiles while DeepSD is relatively stable. Though RMSE, Corr, and Skill becomes worse at these extremes, DeepSD

consistently outperforms BCSD, most often with thinner confidence bounds. These results show DeepSD’s ability to perform well for increasingly extreme precipitation events. DeepSD’s performance is impressive given that literature has shown that traditional techniques tend to fail when downscaling averages and extremes simultaneously. We hypothesize that capturing nearby spatial information allows DeepSD to isolate areas where extreme precipitation events are more likely than others.

6.3 Scalability on NASA’s NEX

Comprehensive studies of climate change requires much more than a single ESM simulation but rather multiple projections from different models, emission scenarios, and initial conditions in order to capture uncertainty. In total, CMIP5 contains more than 20 models at 4 emission scenarios (Representative Concentration Pathways (RCPs) 2.6,4.5,6.5,8.6), a variable number of initial conditions, and multiple climate variables at a daily temporal resolution. Generally, each prospective projection is available from 2006 to 2100 while retrospective projections are available from 1850 to 2005. Limiting the downscaled projections to encompass CONUS at $1/8^\circ \times 1/8^\circ$, a single simulation requires 134MB. Following the current timeframes of downscaled projections on NASA’s NEX platform, downscaling from the year 1950 to 2005 requires 7.4GB of storage while each prospective run needs 13GB. Hence, the final dataset size is 1.2TB. When downscaled further to PRISM’s native resolution, $1/16^\circ \times 1/16^\circ$, the dataset size increases to approximately 5TB. Furthermore, the dataset scales linearly as more variables are added, including temperature minimum and maximum.

We test computational scalability by computing the amount of time taken to downscale 1 year of CONUS, presented in Table 1. For DeepSD, this includes the 3 feed-forward processes and their corresponding interpolations computed on a single GPU in NVIDIA DIGITS DevBox. Runtime for each of the ASD methods is estimated and scaled from the length of time to downscale the 1000 selected locations using 40 CPUs in parallel (Intel Xeon CPU E5-2680 2.8GHz). BCSD’s runtime is computed as the amount of time taken to quantile map, interpolate, and scale a years worth of data. We find that BCSD and DeepSD widely outperform the ASD approaches. While BCSD provides the quickest runtime, DeepSD is still scalable.

Though DeepSD is a highly scalable method, due to a single feed-forward neural network architecture, generating such large datasets still requires heavy computational power. However, storage and compute resources are satisfied by dedicated access to the Pleiades supercomputer housed in NASA’s Advanced Supercomputer Division (HECC) at NASA Ames. High resolution projections can be quickly computed using GPU’s, which are available on each node, in coordination with the Message Passing Interface (MPI). High-resolution projections are then stored on NEX’s filesystem which has currently 2.3PB of rapid-access storage in addition to large scale tape storage accessible on the HECC platform. In this paper we present a methodology for statistical downscaling, DeepSD, that leverages recent advances in image super-resolution and convolutional neural networks. DeepSD differs from previous SD methods by explicitly capturing spatial structure while improving scalability. A brief comparison with baseline SD techniques, BCSD and ASD,

shows promising results in predictive capabilities when downscaling precipitation over the continental United States. Lastly, we describe how DeepSD can be scaled using NASA's Earth Exchange platform to provide an ensemble of downscaled climate projections from more than 20 ESMs.

7 CONCLUSION

Though DeepSD shows promise for SD, there are still some limitations in our experimentation regarding spatial and temporal generalization. An advantage of DeepSD is that a single trained model is able to downscale spatial heterogeneous regions. However, we do not test predictability in regions where the model was not trained. Future work will examine this hypothesis to understand DeepSD's credibility in regions with few observations. Second, we do not test temporal non-stationarity, a longstanding problem in statistical downscaling. Evaluation under non-stationarity can be tested using approaches presented by Salvi et al. [29], such that training and testing data is split between cold/warm years. As there is a single model for all locations, including cold and warm climates, we hypothesize that DeepSD is capable of capturing non-stationarity.

Furthermore, future work can improve multiple facets of DeepSD. For instance, the inclusion of more variables such as temperature, wind, and humidity at different pressure levels of the atmosphere may capture more climate patterns. Also, downscaling multiple climate variables simultaneously could be explored to find similar spatial patterns in the high-resolution datasets, such as high temperatures and increased precipitation. Most importantly, DeepSD fails to capture uncertainty around its downscaled projections, a key factor in adapting to climate change. Recent advances in Bayesian Deep Learning concepts [12] may aid in quantifying uncertainty. Though these limitations exist, DeepSD is a scalable architecture with high predictive capabilities which provides a novel framework for statistical downscaling.

ACKNOWLEDGMENTS

This work was supported by NASA Earth Exchange (NEX), National Science Foundation CISE Expeditions in Computing under grant number: 1029711, National Science Foundation CyberSEES under grant number: 1442728, and National Science Foundation BIGDATA under grant number: 1447587. The GTOPO30 dataset was distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS/EROS, Sioux Falls, SD. <http://lpdaac.usgs.gov>. We thank Arindam Banerjee for valuable comments.

REFERENCES

- [1] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, and others. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016).
- [2] Saikat Basu, Sangram Ganguly, Supratik Mukhopadhyay, Robert DiBiano, Manohar Karki, and Ramakrishna Nemani. 2015. DeepSat: a learning framework for satellite imagery. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 37.
- [3] G Bürger. 1996. Expanded downscaling for generating local weather scenarios. *Climate Research* 7, 2 (1996), 111–128.
- [4] G. Bürger, T. Q. Murock, a. T. Werner, S. R. Sobie, and a. J. Cannon. 2012. Downscaling extremes-an intercomparison of multiple statistical methods for present climate. *Journal of Climate* 25, 12 (June 2012), 4366–4388. <http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00408.1>
- [5] William R Burrows, Mario Benjamin, Stephen Beauchamp, Edward R Lord, Douglas McCollor, and Bruce Thomson. 1995. CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada. *Journal of applied meteorology* 34, 8 (1995), 1848–1862.
- [6] Alex J Cannon. 2011. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & geosciences* 37, 9 (2011), 1277–1284.
- [7] Hong Chang, Dit-Yan Yeung, and Yimin Xiong. 2004. Super-resolution through neighbor embedding. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 1. IEEE, I–I.
- [8] Christopher Daly, Michael Halbleib, Joseph I Smith, Wayne P Gibson, Matthew K Doggett, George H Taylor, Jan Curtis, and Phillip P Pasteris. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International journal of climatology* 28, 15 (2008), 2031–2064.
- [9] Christopher Daly, Ronald P Neilson, and Donald L Phillips. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of applied meteorology* 33, 2 (1994), 140–158.
- [10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2014. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*. Springer, 184–199.
- [11] William T Freeman, Thouis R Jones, and Egon C Pasztor. 2002. Example-based super-resolution. *IEEE Computer graphics and Applications* 22, 2 (2002), 56–65.
- [12] Yariv Gal. 2016. *Uncertainty in Deep Learning*. Ph.D. Dissertation. Ph. D. thesis, University of Cambridge.
- [13] Subimal Ghosh. 2010. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. *Journal of Geophysical Research: Atmospheres* 115, D22 (2010).
- [14] Daniel Glasner, Shai Bagon, and Michal Irani. 2009. Super-resolution from a single image. In *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 349–356.
- [15] Muhammad Zia Hashmi, Asaad Y Shamseldin, and Bruce W Melville. 2011. Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. *Stochastic Environmental Research and Risk Assessment* 25, 4 (2011), 475–484.
- [16] Masoud Hessami, Philippe Gachon, Taha BMJ Ouarda, and André St-Hilaire. 2008. Automated regression-based statistical downscaling tool. *Environmental Modelling & Software* 23, 6 (2008), 813–834.
- [17] Hugo Hidalgo, Michael Dettinger, and Daniel Cayan. 2008. Downscaling with Constructed Analogues: Daily Precipitation and Temperature Fields over the United States. (2008).
- [18] Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* 313, 5786 (2006), 504–507.
- [19] Richard A Kerr. 2013. Forecasting regional climate change flunks its first test. *Science* 339, 6120 (2013), 638–638.
- [20] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [21] Elizabeth C Mannshardt-Shamseldin, Richard L Smith, Stephan R Sain, Linda O Mearns, Daniel Cooley, and others. 2010. Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *The Annals of Applied Statistics* 4, 1 (2010), 484–502.
- [22] Alon Manor and Sigalit Berkovic. 2015. Bayesian Inference aided analog downscaling for near-surface winds in complex terrain. *Atmospheric Research* 164 (2015), 27–36.
- [23] Edwin P Maurer and Hugo G Hidalgo. 2008. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. (2008).
- [24] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 807–814.
- [25] Robert J Nicholls and Anny Cazenave. 2010. Sea-level rise and its impact on coastal zones. *science* 328, 5985 (2010), 1517–1520.
- [26] Rajendra K Pachauri, MR Allen, VR Barros, J Broome, W Cramer, R Christ, JA Church, L Clarke, Q Dahe, P Dasgupta, and others. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (2014).
- [27] SE Perkins, AJ Pitman, NJ Holbrook, and J McAneney. 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *Journal of climate* 20, 17 (2007), 4356–4376.
- [28] David W Pierce, Daniel R Cayan, and Bridget L Thrasher. 2014. Statistical Downscaling Using Localized Constructed Analogs (LOCA)*. *Journal of Hydrometeorology* 15, 6 (2014), 2558–2585.
- [29] Kaustubh Salvi, Subimal Ghosh, and Auroop R. Ganguly. 2015. Credibility of statistical downscaling under nonstationary climate. *Climate Dynamics* May 2014 (2015). DOI: <http://dx.doi.org/10.1007/s00382-015-2688-9>

- [30] Quirin Schiermeier. 2010. The real holes in climate science. *Nature News* 463, 7279 (2010), 284–287.
- [31] Samuel Schuster, Christian Leistner, and Horst Bischof. 2015. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3791–3799.
- [32] Anne MK Stoner, Katharine Hayhoe, Xiaohui Yang, and Donald J Wuebbles. 2013. An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology* 33, 11 (2013), 2473–2494.
- [33] Karl E Taylor, Ronald J Stouffer, and Gerald A Meehl. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* 93, 4 (2012), 485.
- [34] Bridget Thrasher, Edwin P Maurer, C McKellar, and PB Duffy. 2012. Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences* 16, 9 (2012), 3309–3314.
- [35] Radu Timofte, Vincent De Smet, and Luc Van Gool. 2014. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*. Springer, 111–126.
- [36] Kevin E Trenberth. 2012. Framing the way to relate climate extremes to climate change. *Climatic change* 115 (2012), 283–290.
- [37] Thomas Vandal, Evan Kodra, and Auroop R Ganguly. 2017. Intercomparison of Machine Learning Methods for Statistical Downscaling: The Case of Daily and Extreme Precipitation. *arXiv preprint arXiv:1702.04018* (2017).
- [38] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research* 11, Dec (2010), 3371–3408.
- [39] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. 2015. Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE International Conference on Computer Vision*. 370–378.
- [40] RL Wilby, SP Charles, E Zorita, B Timbal, P Whetton, and LO Mearns. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. (2004).
- [41] Andrew W Wood, Lai R Leung, V Sridhar, and DP Lettenmaier. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic change* 62, 1-3 (2004), 189–216.
- [42] Jianchao Yang, Zhe Lin, and Scott Cohen. 2013. Fast image super-resolution based on in-place example regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1059–1066.
- [43] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. 2010. Image super-resolution via sparse representation. *IEEE transactions on image processing* 19, 11 (2010), 2861–2873.
- [44] Liangpei Zhang, Lefei Zhang, and Bo Du. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine* 4, 2 (2016), 22–40.
- [45] Xianliang Zhang and Xiaodong Yan. 2015. A new statistical precipitation downscaling method with Bayesian model averaging: a case study in China. *Climate Dynamics* 45, 9-10 (2015), 2541–2555.