

Generating High Resolution Climate Change Projections through Single Image Super-Resolution: An Abridged Version

Thomas Vandal¹, Evan Kodra², Sangram Ganguly^{3,4},
Andrew Michaelis⁵, Ramakrishna Nemani^{3,6} and Auroop R. Ganguly^{1,2}

¹ Northeastern University, Civil and Environmental Engineering

² risQ Inc.

³ NASA Ames Research Center

⁴ Bay Area Environmental Research Institute

⁵ University Corporation, Monterey Bay

⁶ NASA Advanced Supercomputing Division

vandal.t@husky.neu.edu, evan.kodra@risq.io, sangram.ganguly@nasa.gov, amac@hyperplane.org,
rama.nemani@nasa.gov, 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. 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 with multi-scale input channels for statistical downscaling of climate variables. A comparison of DeepSD to four state-of-the-art methods 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.

extreme weather events with increasing intensity, duration, and frequency [Trenberth, 2012]. 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 [Taylor *et al.*, 2012]. 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 [Schiermeier, 2010].

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 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 [Hessami *et al.*, 2008], neural networks [Cannon, 2011], and support vector machines [Ghosh, 2010], have been applied to SD. 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.

1 Introduction

Climate change is causing detrimental effects to society's well being as temperatures increase, extreme events become more intense [Pachauri *et al.*, 2014], and sea levels rise [Nicholls and Cazenave, 2010]. 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

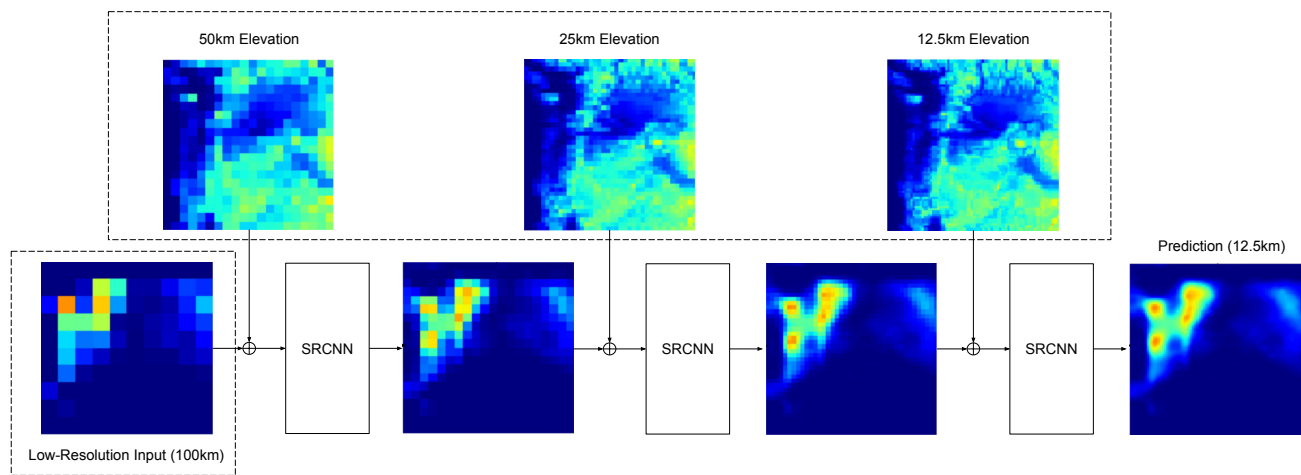


Figure 1: (Best viewed in color) **DeepSD Framework**: This end-to-end framework of DeepSD includes a set of stacked SRCNNs where the inputs (dashed boxes) include a low-resolution input and high-resolution auxiliary variables.

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. Advances in single image super-resolution (SR) correspond well to SD, which learns a mapping between low- and high-resolution images. Through experimentation, we found that super-resolution convolutional neural networks were able to capture spatial information in climate data to improve beyond existing methods.

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.
- 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.

2 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. 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.

3 Methods

3.1 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. [Dong *et al.*, 2014], 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 $\mathbf{X}$ while the HR label is denoted as $\mathbf{Y}$. A three layer CNN, F , with rectified linear unit activations can be defined such that $F(\mathbf{X})$ attempts to reconstruct $\mathbf{Y}$. A Euclidean loss function is used to the parameters (weights and biases) of F to define

an optimization objective as:

$$\operatorname{argmin}_{\Theta} \sum_{i=1}^n \|F(\mathbf{X}_i; \Theta) - \mathbf{Y}_i\|_2^2 \quad (1)$$

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

3.2 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 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 [Vincent *et al.*, 2010] and stacked RBMs for deep belief networks [Hinton and Salakhutdinov, 2006]. 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.

3.3 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 it's associated input/output pairs. As presented in figure 1, inference is executed by starting with the lowest resolution image with it's associated HR elevation to predict the first resolution enhancement. The next resolution enhancement is estimated from the previous layer's estimate and it's associated HR elevation. This process is repeated for each trained SRCNN. Figure 1 illustrates this process with a precipitation event and it's various resolution improvements. We see that this stacked process allows the model to capture both regional and local patterns.

4 Experiments

4.1 Data

As ESM outputs do not provide information at the high resolution scales, we must use observational datasets to learn our SD model. Hence, in this context, a gridded observation dataset acts as a proxy which can then be applied to

ESMs after training. In our experiments, 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 [Daly *et al.*, 2008]. 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 [Pierce *et al.*, 2014]. For topography we use the the GTOPO30 elevation (30 arcsec spatial resolution) dataset distributed by the Land Processes Distributed Active Archive Center (LP DAAC). The goal is then to learn a mapping between our low-resolution and high-resolution datasets.

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 it's 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.

4.2 Training Parameters

All SRCNNs are trained with the same set of parameters, selected using those found to work well by Dong *et al.* [Dong *et al.*, 2014]. 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. 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 [Kingma and Ba, 2014] 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 [Abadi *et al.*, 2016] 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.

4.3 Results

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 2 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 [Perkins *et*

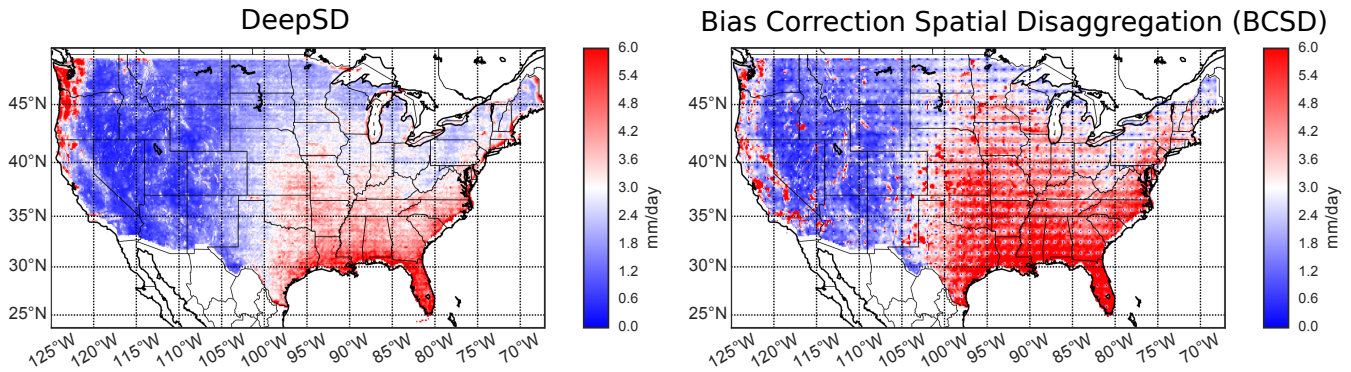


Figure 2: (Best viewed in color) Daily Root Mean Square Error (RMSE) computed at each location for years 2006 to 2014 (test set) in CONUS for Left) DeepSD and Right) BCSD. Red corresponds to high RMSE while blue corresponds to low RMSE.

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	—
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.

et al., 2007], is used to measure distribution similarity between 0 and 1 where 1 corresponds to identical distributions.

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 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 [Vandal *et al.*, 2017a], 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. 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 2 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. We encourage the reader to refer to [Vandal *et al.*, 2017b] for more thorough results and DeepSD’s application on the NASA Earth Exchange (NEX) platform.

5 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.* [Salvi *et al.*, 2015], 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 [Gal, 2016; Vandal *et al.*, 2018] 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

- [Abadi *et al.*, 2016] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [Cannon, 2011] Alex J Cannon. Quantile regression neural networks: Implementation in r and application to precipitation downscaling. *Computers & geosciences*, 37(9):1277–1284, 2011.
- [Daly *et al.*, 2008] Christopher Daly, Michael Halbleib, Joseph I Smith, Wayne P Gibson, Matthew K Doggett, George H Taylor, Jan Curtis, and Phillip P Pasteris. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states. *International journal of climatology*, 28(15):2031–2064, 2008.
- [Dong *et al.*, 2014] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199. Springer, 2014.
- [Gal, 2016] Yarin Gal. *Uncertainty in Deep Learning*. PhD thesis, Ph. D. thesis, University of Cambridge, 2016.
- [Ghosh, 2010] Subimal Ghosh. Svm-pgsl coupled approach for statistical downscaling to predict rainfall from gcm output. *Journal of Geophysical Research: Atmospheres*, 115(D22), 2010.
- [Hessami *et al.*, 2008] Masoud Hessami, Philippe Gachon, Taha BMJ Ouada, and André St-Hilaire. Automated regression-based statistical downscaling tool. *Environmental Modelling & Software*, 23(6):813–834, 2008.
- [Hinton and Salakhutdinov, 2006] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [Kingma and Ba, 2014] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [Nicholls and Cazenave, 2010] Robert J Nicholls and Anny Cazenave. Sea-level rise and its impact on coastal zones. *science*, 328(5985):1517–1520, 2010.
- [Pachauri *et al.*, 2014] Rajendra K Pachauri, MR Allen, VR Barros, J Broome, W Cramer, R Christ, JA Church, L Clarke, Q Dahe, P Dasgupta, et al. 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.
- [Perkins *et al.*, 2007] SE Perkins, AJ Pitman, NJ Holbrook, and J McAneney. 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):4356–4376, 2007.
- [Pierce *et al.*, 2014] David W Pierce, Daniel R Cayan, and Bridget L Thrasher. Statistical downscaling using localized constructed analogs (loca)*. *Journal of Hydrometeorology*, 15(6):2558–2585, 2014.
- [Salvi *et al.*, 2015] Kaustubh Salvi, Subimal Ghosh, and Auroop R. Ganguly. Credibility of statistical downscaling under nonstationary climate. *Climate Dynamics*, (May 2014), 2015.
- [Schiermeier, 2010] Quirin Schiermeier. The real holes in climate science. *Nature News*, 463(7279):284–287, 2010.
- [Taylor *et al.*, 2012] Karl E Taylor, Ronald J Stouffer, and Gerald A Meehl. An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485, 2012.
- [Trenberth, 2012] Kevin E Trenberth. Framing the way to relate climate extremes to climate change. *Climatic change*, 115:283–290, 2012.
- [Vandal *et al.*, 2017a] Thomas Vandal, Evan Kodra, and Auroop R Ganguly. Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation. *arXiv preprint arXiv:1702.04018*, 2017.
- [Vandal *et al.*, 2017b] Thomas Vandal, Evan Kodra, Sangram Ganguly, Andrew Michaelis, Ramakrishna Nemani, and Auroop R Ganguly. Deepsd: Generating high resolution climate change projections through single image super-resolution. In *23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2017.
- [Vandal *et al.*, 2018] Thomas Vandal, Evan Kodra, Jennifer Dy, Sangram Ganguly, Ramakrishna Nemani, and Auroop R Ganguly. Quantifying uncertainty in discrete-continuous and skewed data with bayesian deep learning. *arXiv preprint arXiv:1802.04742*, 2018.
- [Vincent *et al.*, 2010] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.