Conditioned Spatial Downscaling of Climate Variables

Alex Ling Yu Hung*
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
alexhung96@ucla.edu

Evan Becker*
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
evbecker@ucla.edu

Ted Zadouri*
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
tedzadouri@g.ucla.edu

Aditya Grover
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
adityag@cs.ucla.edu

Abstract

Global Climate Models (GCM) play a vital role in assessing the large-scale impacts of climate change. Downscaling methods can translate coarse-resolution climate information from GCM to high-resolution predictions to forecast regional effects. Unfortunately, current downscaling methods struggle to fully take into account spatial relationships among variables, especially at long distances. In this work, we propose an instance-conditional pixel synthesis generative adversarial network (ICPS-GAN), wherein conditioning on spatial information is an explicit way of providing the GAN with previous high-resolution and current low-resolution data, resulting in an enhancement of the general performance. Experimental results on precipitation forecast for US region data outperform both traditional and other learning-based methods when extrapolating in space. The code is available at https://github.com/evbecker/climate-spatial-downscaling

1 Introduction

Climate change has immense impacts on people's lives and ecosystems across the world, and with the exacerbation of global warming, there are more severe weather conditions now than ever before (Nicholls & Cazenave, 2010; Villén-Peréz et al., 2020; Wang et al., 2020; Giorgi et al., 2019). Correctly predicting climate change can save people a lot of time, money, resources, and even lives (Chakraborty et al., 2000; Guo et al., 2018), as people will be able to take necessary precautions before some unexpected conditions based on the prediction. However, predicting such changes in the climate is not an easy task to do. In recent years, researchers have attempted to apply learning-based methods to climate problems (Ardabili et al., 2019; Kareem et al., 2021), since deep learning-based approaches are known for discovering complex underlying patterns.

Global Climate Models (GCMs) are effective in predicting general climate conditions, and estimates based on these models have been applied in many fields, including Earth Science, engineering, economics, and risk analysis (Groenke et al., 2020). However, GCMs usually operate at a lower resolution, having a limited ability to provide regional or fine-detailed predictions, even though they are useful in analyzing the general trend globally. Moreover, there are more large-scale, lower-resolution data available than lower-scale, higher-resolution data, which makes building high-

*Equal contribution

NeurIPS 2022 AI for Science Workshop.

resolution models more difficult. Therefore, generating higher resolution climate projections is important for precise local climate analysis.

Statistical downscaling is the process of using some global information to make local predictions. Since global climate data is easier to access, statistical downscaling is a good way to make use of the low-resolution data for high-resolution climate predictions. The statistical downscaling builds the relationship between the local climate variables and global predictors and uses such a relationship to make future predictions of the local climate variables based on the global predictors (Hoar & Nychka, 2008). In this work, we build a generative deep statistical downscaling model for local climate variable predictions.

Our contributions are as follows:

1. We propose a type of instance-conditional generative adversarial network (GAN) that addresses the issue of long range spatial dependencies by explicitly conditioning on location data and previous time-steps high-resolution images.
2. We characterize this model's performance on downscaling tasks between real-world US precipitation models at different resolutions, and demonstrate its improvement over standard methods.

2 Problem Statement

Many current statistical downscaling methods formulate the downscaling problem as finding some direct mapping function from low resolution to high resolution data (e.g. some pointwise regression). The caveat with such methods is that they lack any underlying distribution in the high resolution space (which can be helpful for computing uncertainties associated with downscaled predictions). Instead, we pose the super-resolution task as a problem of finding conditional distributions.

Let $I_{HR}^{(t)} \in \mathbb{R}^{(H\times W\times C)}$ and $I_{LR}^{(t)} \in \mathbb{R}^{(H\times W\times C)}$ be a sequence of random variables representing high-resolution (HR) and low-resolution (LR) images, respectively with height H, width W and depth C. Each image consists of one or more channels of weather data at set latitudes and longitudes (x,y) for various time steps $t \in T$. Our goal is to learn a generator that has the following mapping:

$$G : (x, y, t, I_{HR}^{(t')}, I_{LR}^{(t)}, \mathbf{z}) \longrightarrow I_{HR}^{(t)} \tag{1}$$

Here, $\mathbf{z}$ is some latent variable following a tractable distribution (e.g. normal). In this paper, the objective is to estimate the high resolution image $I_{HR}^{(t)}$ given a previous high resolution image ($I_{HR}^{(t')}$ for some $t' < t$) and a current low resolution image ($I_{LR}^{(t)}$). The following tuple is the training input $(I_{HR}^{(t)}, I_{HR}^{(t')}, I_{LR}^{(t)})$. For inference, the model has access to $I_{LR}^{(t)}$ to make the prediction where that particular time-step has not been part of the training. The metrics that will be utilized to assess model performance over the evaluation period and quantify the representation of mean properties will be based on those reported for the VALUE baseline ensemble (Widmann et al., 2019) such as root mean square error (RMSE) and Pearson correlation.

3 Related Works

3.1 Traditional Methods

In traditional methods, the spatial structure of daily values for climate impacts is an important factor to consider. Easterling (1999); Kettle & Thompson (2004); Huth et al. (2008) discuss the correlations between time series at different spatial locations via Perfect Prog (PP) methods and find that large-scale predictors tend to overestimate spatial correlations, while more local methods underestimate them. PP methods work the well in general, but the performance drops in summer and winter Ayar et al. (2016); Huth et al. (2015). In addition, stochastic PP methods are useful when they explicitly model spatial structures. Cannon (2008); Wilks (2012); Hu et al. (2013) find that both Nonhomogeneous Hidden Markov Models (NHMMs) and Generalized Linear Models (GLMs) are likely to underestimate the correlation, while NHMMs are better under certain circumstances.

Analog methods search for historical data to match the current or forecast local variables such as precipitation over an area to similar times in the past. Pierce et al. (2014) performs the statistical downscaling using Localized Constructed Analogs (LOCA). Gutmann et al. (2014) compare several analog methods, and most methods overestimate the spatial correlation, except for the method combining bias correction for monthly fields and spatial disaggregation (BCSDm). They think the reason why BCSDm does not overestimate the spatial correlation is that the method inherits the spatial variability from the observations, rather than from the driving model.

3.2 Learning-based Methods

Methods based on deep learning have also been used to solve statistical downscaling. Sun & Lan (2021) use a simple Convolutional Neural Network (CNN) to predict the daily temperature and precipitation in China. Babaousmail et al. (2021) apply a convolutional autoencoder on data from North Africa where the low-resolution global data is the input and the high-resolution regional data is the output. Misra et al. (2018) propose a Recurrent Neural Network (RNN) model with Long Short-Term Memory (LSTM) to capture the spatial and temporal dependencies in local rainfall. The study tests their model on two datasets: precipitation in the Mahanadi basin in India and precipitation in the Campbell River basin in Canada, instead of the GCMs. Sekiyama (2020) treat the climate data statistical downscaling problem as a single-image super-resolution (SR) problem.

Statistical downscaling algorithms that are based on deep learning have been shown to have better empirical accuracy than traditional methods, but they are mostly like black box (Jebeile et al., 2021). Accarino et al. (2021) uses a multi-scale Generative Adversarial Network (GAN) to downscale temperature fields and Chaudhuri & Robertson (2020) proposes CliGAN, a Wasserstein GAN that is used to downscale large-scale annual maximum precipitation given by simulation of multiple atmosphere-ocean global climate models. Leinonen et al. (2020) develops a recurrent, stochastic super-resolution GAN that can generate ensembles of time-evolving high-resolution atmospheric fields given an input sequence of low-resolution images of the same field. Vaughan et al. encodes the low-resolution predictor to a latent variable and retrieves the local climate variable as a query of longitude, latitude, and elevation via the latent neural process utilizing ConvLNP (Vaughan et al., 2022).

In terms of generative model-based Super Resolution (SR) algorithms, Ledig et al. (2017) proposes SRGAN which is a GAN-based network optimized for a new perceptual loss. ESRGAN (Wang et al., 2018) includes a Residual-in-Residual Dense Block (RRDB) to combine multi-level residual networks and dense connections. ESRGAN+ (Rakotonirina & Rasoanaivo, 2020) adds Gaussian noise vectors to ESRGAN after each residual along with a learned scaling factor. In GLEAN (Chan et al., 2021), to make the model more expressive, each block of the generator takes in a latent vector while also being conditioned on the output of the encoder. Hyun & Heo (2020) proposes VarSR-Net, which is a variational inference-base SR algorithm. Lugmayr et al. (2020) proposes SRFlow and uses normalizing flow models for the SR problem. Based on SRFlow, Jo et al. (2021) include more convolutional layers to have a large receptive field in a single flow step for better results.

4 Method

4.1 Pixel Synthesis

Inspired by the recent work of He et al. (2021); Anokhin et al. (2020), we leverage the conditionally independent pixel synthesis generator architecture whose generative model computes climate values at individual locations, independent of others, given a random noise vector, low-resolution weather embedding, and positional embedding of its coordinates in time and space. In addition to the latitude, longitude, and time used by He et al. (2015) to generate RGB satellite images, we also encode elevation information. Elevation is an important spatial feature that should allow the model to interpolate weather data more accurately across mountainous terrain. To be specific, we encode the latitude, longitude, elevation, and time stamp through a learnable embedding function f_γ, parameterized by γ, s.t.

$$v = f_\gamma(x, y, h, t) \tag{2}$$

where v is the embedded positional feature, x, y, h and t is the latitude, longitude, elevation, and time stamp respectively. There is also a function f_θ, which takes in a low-resolution image at the current time step and a high-resolution image from the previous time step and outputs the encoded image features m:

$$m = f_\theta(I_{lr}^{(t)}, I_{hr}^{(t')}) \tag{3}$$

where $I_{lr}^{(t)}$ is the low-resolution image at time t and $I_{hr}^{(t')}$ is the high-resolution image at time t'. The generator G takes in the encoded image feature m, positional feature v, and a noise vector z to generate the predicted high-resolution image ${I'_{hr}}^{(t)}$:

$${I'_{hr}}^{(t)} = G(z, v, m) = G(X, \mathbf{z}|I_{lr}^{(t)}, I_{hr}^{(t')}) \tag{4}$$

where $X = (x, y, h, t)$.

4.2 Instance-Conditioned Loss Function

In the given work, we propose an instance-conditional pixel synthesis generative adversarial network (ICPS-GAN) where we utilize the super-resolution architecture via conditional pixel synthesis (He et al., 2021) but with an alteration to its loss function. Instead of training the generator with a combination of the conditional GAN loss with L0 loss, we use Instance-Conditioned GAN (IC-GAN) (Casanova et al., 2021):

$$G^* = \mathbb{L}_{(IC-GAN)}(G, D) + \lambda \mathbb{L}_{L1}(G) \tag{5}$$

$$\begin{aligned} \mathbb{L}_{(IC-GAN)}(G, D) =& \mathbb{E}[\log D(I_{hr}^{(t)}, X, I_{lr}^{(t)}, I_{hr}^{(t')})] + \\ & \mathbb{E}[1 - \log D(G(X, \mathbf{z}|I_{lr}^{(t)}, I_{hr}^{(t')}, f(x_i)), X, I_{lr}^{(t)}, I_{hr}^{(t')})] \end{aligned} \tag{6}$$

where X in equation 6 is the temporal-spatial coordinate grid (x, y, h, t), while G and D in equation 5 and 6 represent the Generative and Discriminator networks respectively, and $\mathbf{z} \in \mathbb{R}^z$ is the noise vector. The main objective of the IC-GAN is to learn the distribution around each datapoint. Essentially, leveraging the overlapping clusters in the data manifold where each cluster is represented by a datapoint or an instance x_i. The underlying distribution $p(x)$ of the data will be a mixture of conditional distributions $p(x|f(x_i))$ where f is an embedding function for each instance. While the original work pre-trains this embedding function on either classification or self-supervised learning tasks, we chose to incorporate this training directly into the generator updates, with the hypothesis that we will still learn encodings useful for clustering low-resolution inputs.

5 Datasets

We train and evaluate our downscaling method on daily weather data from a custom combination of ERA-iterim, CPC, and WRF datasets. Our base datasets all overlap temporally for a period of 7 years (2000-2006), giving us a minimum of 2,282-time samples for each type of measurement, while ERA-iterim and CPC overlap temporally from 2000 to 2018. At each time point, measurements for a single variable will consist of a 2D image whose pixels correspond to a specific latitude and longitude coordinate pair within the United States. For this work, we focus on daily accumulated precipitation, measured in millimeters (mm), which provides the opportunity for our model to capture complex long-range spatial and temporal dependencies across the data.

The low-resolution ERA-iterim dataset has a native resolution of approximately 0.75×0.75 degrees and contains global hourly measurements for weather variables such as temperature, pressure, and precipitation (Berrisford et al., 2011). The higher resolution CPC US Unified Precipitation data is provided by NOAA. Resolution is 0.25×0.25 degrees and spans the continental United States (CONUS).The WRF dataset contains very high resolution data at approximately $\frac{1}{32} \times \frac{1}{32}$ degrees (Rasmussen & Liu, 2017). Total precipitation (mm) is recorded at hourly intervals and preprocessed into daily intervals. Additionally, data is stored not in a regular latitude and longitude grid but in a projected Lambert Conformal grid. As shown in figure 1, we crop five different 10×10 degree squares across the United States to create climate "images" that we can then feed into our network. We discuss the train and test splits of these images later in the section 6.2. We hope that aligning these three datasets at different resolutions will provide a flexible and reusable benchmark for future research. Additional prepossessing and feature descriptions are provided in the appendix section A.1.

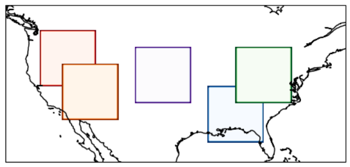


Figure 1: Regional 10x10 degree "patches" of data are taken and used to train the downscaling model: Northwestern (red), Northeastern(green), Southwestern (orange), Southeastern (blue) regions are shown.

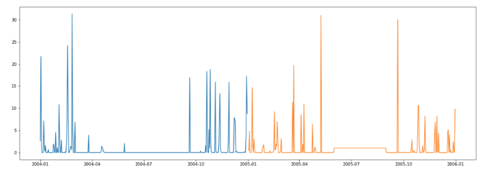


Figure 2: High-Resolution training data for the precipitation (blue) year 2004 vs. High-Resolution prediction (orange) year 2005 for a single coordinate.

6 Experiments and Results

6.1 Baseline Methods

BCSD For the bias correction and spatial disaggregation (BCSD) baseline, the quantile mapping from the low-resolution to high-resolution data is fit for each day using the training data. In Figure 2, for a fixed longitude and latitude, the precipitation is projected for the year 2005, demonstrated in orange, where the model was trained on the data represented in blue from the previous year.

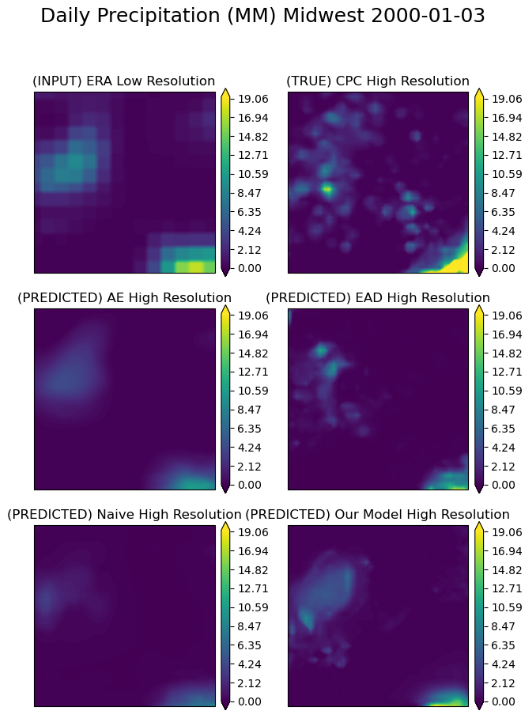
Vanilla Autoencoder For the Vanilla Autoencoder (Vanilla AE), we follow the network architecture described below. The encoder side consists of 3 layers, where each layer has two 3×3 convolutional layers. Each convolutional layer is followed by a Leaky Rectified Linear Unit (ReLU) and batch normalization. At the end of each layer, we downsample the feature map by max pooling. There are 16, 32, and 64 filters in convolutional layers in each of the three layers, respectively. The decoder follows the opposite structure except that at the beginning of each layer, we upsample the feature map by bilinear interpolation.

Naive Version of EAD In this baseline approach, we implemented a naive version of EAD (He et al., 2021), where we use the same architecture as the generator of the GAN but without the noise vector, coordinates, high-resolution image from the previous time step, and the discriminator. Therefore, the naive version of EAD (Naive EAD) is some form of the deterministic autoencoder.

EAD For the last baseline comparison, we re-implement the EAD method by He et al. (2021), where it takes in the spatial coordinate as well as the time stamp as an input to the network, which allows it to locate the exact spatial and temporal information of the given image.

6.2 Implementation Details

For the experiments in this section, we held out the data from Northwestern, Northeastern, Southwestern, and Southeastern from 2001 as the validation set and the ones from the Midwest from the years 2001 to 2007 as a separate test of an unseen region. Due to computational limitations, the networks in this work were trained on the data from the 4 regions (Northwestern, Northeastern, Southwestern, and Southeastern) from 2000 to 2007. In Section 6.7, the networks used in the experiments were trained on the data from those 4 regions from 2000 to 2018. In Section 6.5, the training was only done in 1 region (Northeastern) from 2002 to 2007. In Section 6.8, we used the CPC data as the low-resolution data with the WRF data as the high-resolution data where we resized the images to 160×160, while in all the other sections in the experiments, we used ERA data as the low-resolution data and CPC data as the high-resolution data where we resized the images to 40×40. For a fair comparison, the training objective for both of the deterministic methods is L1 loss. For the GAN-based methods, we set $\lambda = 0.2$, since this is the optimal value we observe on the validation set, which is in line with the results in Section 6.5 on the test set.



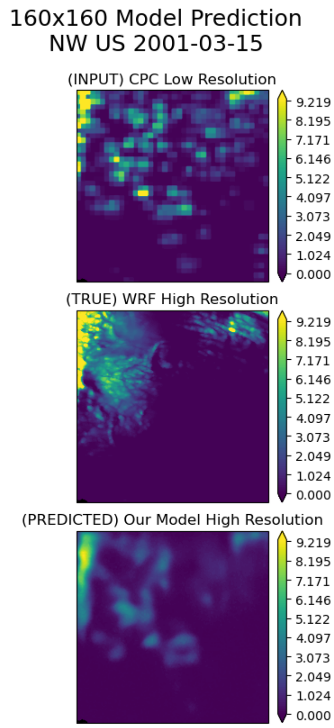


Figure 3: Downscaling predictions for an unknown year. Left column: input low-resolution ERA data, vanilla AE predicted result, naive EAD predicted result. Right column: ground truth CPC high-resolution data, EAD predicted result, our predicted result. The deterministic methods fail to generate meaningful results.

Figure 4: Downscaling predictions from the high-resolution model for an unseen date. The top image is the low-resolution image from the CPC dataset , the middle image is the true WRF high-resolution image, and the bottom image is our model's prediction.

6.3 Baseline Comparison

The baseline quantitative comparison of the held out region can be found in Table 1. Our method outperforms the other methods quantitatively. We believe this is because our method can generalize better to the unseen region with the longitude, latitude, and elevation information. The qualitative results are shown in Fig. 3, where the deterministic methods fail to generate anything meaningful.

Table 1: Comparison against baseline approaches on a held out region

	RMSE	Pearson Correlation
BCSD	4.21	0.183
Vanilla AE	2.90 ± 3.28	0.279 ± 0.272
Naive EAD	2.88 ± 3.23	0.296 ± 0.270
EAD	2.98 ± 3.21	0.235 ± 0.248
ours	2.82 ± 3.14	0.311 ± 0.262

6.4 Ablation Study

We also performed an ablation study on the unseen region, which is shown in Table 2. It can be observed that adding both the elevation data and conditioning the GAN on the instance improves the performance of the model while using both improves the performance the most. Adding the elevation information to the network is more useful for unseen regions as it improves the Pearson Correlation by a significant factor, while for unseen years the improvement is marginal. We conclude that if the network has seen the region during training, it is more likely to implicitly figure out the elevation of the region to some extent since the longitude and latitude are explicitly given. Therefore, explicitly telling the network the elevation information allows the network to learn the correlation between precipitation and elevation, which is shown to be useful for unseen regions. Conditioning the GAN on the instance is a more direct way of giving the GAN information about the current low-resolution data and previous high-resolution data compared with normal GAN, and it is shown to be an effective way of improving the general performance.

Table 2: Ablation study on the unseen region

Elevation	IC	RMSE	Pearson Correlation
		2.98 ± 3.21	0.235 ± 0.248
✓		2.84 ± 3.16	0.275 ± 0.265
	✓	2.83 ± 3.13	0.287 ± 0.258
✓	✓	2.82 ± 3.14	0.311 ± 0.262

6.5 Analysis on the Hyperparameter λ

We additionally analyze the effect of the hyperparameter λ in the loss function on the results. The λ parameter controls how much the GAN loss term $\mathbb{L}_{(IC-GAN)(G,D)}$ and L1 loss term $\mathbb{L}_{L_1}(G)$ contribute to the total loss. We want to investigate how this parameter can impact the final results. $\mathbb{L}_{(IC-GAN)(G,D)}$ indicates how much the generated images look like the real images, while $\mathbb{L}_{L_1}(G)$ determines how well the prediction is. We conducted the same experiment 10 times where we set λ as 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10 during each trial. The results are shown in Table 3, where we can observe that the best results occur when $\lambda = 0.2$.

Table 3: Comparison with different values for λ on a held out year

	RMSE	Pearson Correlation
$\lambda = 0.01$	2.91 ± 3.01	0.260 ± 0.236
$\lambda = 0.02$	2.94 ± 3.17	0.258 ± 0.249
$\lambda = 0.05$	2.94 ± 3.13	0.272 ± 0.221
$\lambda = 0.1$	3.03 ± 2.93	0.213 ± 0.221
$\lambda = 0.2$	3.04 ± 2.95	0.274 ± 0.214
$\lambda = 0.5$	2.93 ± 2.95	0.284 ± 0.233
$\lambda = 1$	2.88 ± 2.87	0.287 ± 0.242
$\lambda = 2$	2.79 ± 2.97	0.318 ± 0.238
$\lambda = 5$	2.76 ± 2.80	0.229 ± 0.241
$\lambda = 10$	2.70 ± 2.78	0.330 ± 0.254

6.6 Analysis of Conditional Image Encodings

As discussed in section 4.2, our embedding function differs from Casanova et al. (2021) in that it is not pretrained on a classification task. We ensure that our embedding still learns some representation useful for clustering by doing principal component analysis (PCA) on the encoded conditioning images. By examining figures 5 and 6, it's clear that our embedding contains information about the type of weather and region that a coordinate is experiencing. Expert domain knowledge would most likely be needed to understand whether the encoding is capturing more complex class information such as weather events (e.g. thunderstorms, blizzards, hurricanes).

6.7 Results on the Full Dataset

The results from the model trained on the full dataset are shown in Table 4. However, we observe that the results are worse than the ones from the model trained on the data from 2002–2007. We hypothesize that this is because the data distribution of later years is more different than that of the early years compared to the test set, as the test set for the unseen region is from 2001 to 2007. Another possible explanation is that the model capacity was not large enough to handle a more diverse data distribution, so the performance dropped.

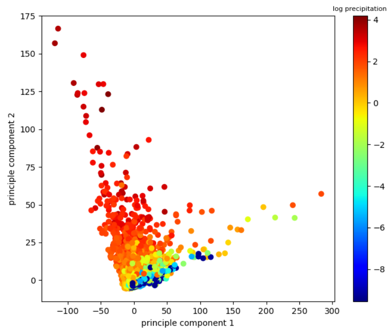


Figure 5: Plotting the first two principal components of our conditioning image embeddings. Each point in the plot corresponds to a specific coordinate and time (from the set of 25 points in the NWUS during the year 2000), while the color corresponds to the log precipitation of the corresponding low resolution image.

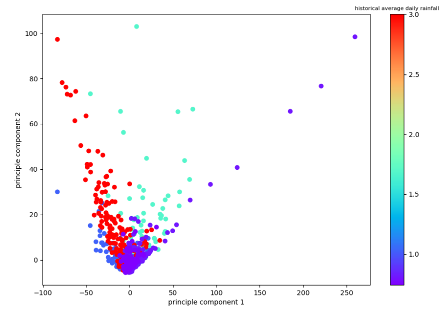


Figure 6: Plotting the first two principal components of our conditioning image embeddings. Each point in the image corresponds to a specific coordinate and time (from the set of 4 points in the NWUS during the year 2000), while the color corresponds to the average daily precipitation for that location.

Table 4: Results on the full dataset

RMSE	Pearson Correlation
2.91 ± 3.16	0.249 ± 0.256

6.8 Generating WRF data based on CPC data

As described in the previous sections, the resolution of ERA data is so poor that it is often impossible to recover some of the higher resolution information from the data. Therefore, we decided to use CPC data as the low-resolution data and WRF data as the high-resolution data in this experiment to test the ability of our method to recover high-resolution information based on data that could provide more information than ERA data. The quantitative results on the unseen regions are shown in Table 5. The qualitative results are shown in Fig. 4, where our prediction looks blurry.

Table 5: Results of generating WRF data from CPC data

RMSE	Pearson Correlation
3.13 ± 3.66	0.243 ± 0.243

7 Conclusion and Future Work

In conclusion, we proposed an instance-conditional pixel synthesis GAN while explicitly feeding in the latitude, longitude, elevation, and time point of each pixel to perform statistic downscaling. The qualitative analysis has shown that our method is superior in generating realistic climate patterns compared with previous methods. The quantitative results have shown that our method has decent generalizability for unseen regions. We believe if our model is trained on more regions than only 4, the performance would be even better compared to the previous methods due to the explicit modeling of the positional information. Due to the constraints of computational resources, we have not been able to perform comprehensive studies on training with the full dataset, generating WRF data based on CPC data, and other climate variables other than precipitation. Future research in these directions would be helpful to exhibit the benefit of our proposed method.

References

Accarino, G., Chiarelli, M., Immorlano, F., Aloisi, V., Gatto, A., and Aloisio, G. Msg-gan-sd: A multi-scale gradients gan for statistical downscaling of 2-meter temperature over the euro-cordex domain. *AI*, 2(4):600–620, 2021.

Anokhin, I., Demochkin, K., Khakhulin, T., Sterkin, G., Lempitsky, V., and Korzhenkov, D. Image generators with conditionally-independent pixel synthesis. *CoRR*, abs/2011.13775, 2020. URL `https://arxiv.org/abs/2011.13775`.

Ardabili, S., Mosavi, A., Dehghani, M., and Várkonyi-Kóczy, A. R. Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In *International Conference on Global Research and Education*, pp. 52–62. Springer, 2019.

Ayar, P. V., Vrac, M., Bastin, S., Carreau, J., Déqué, M., and Gallardo, C. Intercomparison of statistical and dynamical downscaling models under the euro-and med-cordex initiative framework: present climate evaluations. *Climate dynamics*, 46(3-4):1301–1329, 2016.

Babaousmail, H., Hou, R., Gnitou, G. T., and Ayugi, B. Novel statistical downscaling emulator for precipitation projections using deep convolutional autoencoder over northern africa. *Journal of Atmospheric and Solar-Terrestrial Physics*, 218:105614, 2021.

Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, M., Fuentes, M., Kållberg, P., Kobayashi, S., Uppala, S., and Simmons, A. The era-interim archive version 2.0. (1):23, 11 2011. URL `https://www.ecmwf.int/node/8174`.

Cannon, A. J. Probabilistic multisite precipitation downscaling by an expanded bernoulli–gamma density network. *Journal of Hydrometeorology*, 9(6):1284–1300, 2008.

Casanova, A., Careil, M., Verbeek, J., Drozdzal, M., and Romero-Soriano, A. Instance-conditioned GAN. *CoRR*, abs/2109.05070, 2021. URL `https://arxiv.org/abs/2109.05070`.

Chakraborty, S., Tiedemann, A., and Teng, P. S. Climate change: potential impact on plant diseases. *Environmental pollution*, 108(3):317–326, 2000.

Chan, K. C., Wang, X., Xu, X., Gu, J., and Loy, C. C. Glean: Generative latent bank for large-factor image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14245–14254, 2021.

Chaudhuri, C. and Robertson, C. Cligan: A structurally sensitive convolutional neural network model for statistical downscaling of precipitation from multi-model ensembles. *Water*, 12(12):3353, 2020.

Easterling, D. R. Development of regional climate scenarios using a downscaling approach. *Climatic Change*, 41(3):615–634, 1999.

Giorgi, F., Raffaele, F., and Coppola, E. The response of precipitation characteristics to global warming from climate projections. *Earth System Dynamics*, 10(1):73–89, 2019.

Groenke, B., Madaus, L., and Monteleoni, C. Climalign: Unsupervised statistical downscaling of climate variables via normalizing flows. In *Proceedings of the 10th International Conference on Climate Informatics*, pp. 60–66, 2020.

Guo, Y., Gasparrini, A., Li, S., Sera, F., Vicedo-Cabrera, A. M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P. H. N., Lavigne, E., Tawatsupa, B., Punnasiri, K., et al. Quantifying excess deaths related to heatwaves under climate change scenarios: A multicountry time series modelling study. *PLoS medicine*, 15(7):e1002629, 2018.

Gutmann, E., Pruitt, T., Clark, M. P., Brekke, L., Arnold, J. R., Raff, D. A., and Rasmussen, R. M. An intercomparison of statistical downscaling methods used for water resource assessments in the u nited s tates. *Water Resources Research*, 50(9):7167–7186, 2014.

He, K., Zhang, X., Ren, S., and Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9): 1904–1916, 2015.

He, Y., Wang, D., Lai, N., Zhang, W., Meng, C., Burke, M., Lobell, D., and Ermon, S. Spatial-temporal super-resolution of satellite imagery via conditional pixel synthesis. *Advances in Neural Information Processing Systems*, 34, 2021.

Hoar, T. and Nychka, D. Statistical downscaling of the community climate system model (ccsm) monthly temperature and precipitation projections. white paper preprint. *Institute for Mathematics Applied to Geosciences/National Center for Atmospheric Research, Boulder, CO*, 80307, 2008.

Hu, Y., Maskey, S., and Uhlenbrook, S. Downscaling daily precipitation over the yellow river source region in china: a comparison of three statistical downscaling methods. *Theoretical and applied climatology*, 112(3):447–460, 2013.

Huth, R., Kliegrova, S., and Metelka, L. Non-linearity in statistical downscaling: does it bring an improvement for daily temperature in europe? *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 28(4):465–477, 2008.

Huth, R., Mikšovskỳ, J., Štěpánek, P., Belda, M., Farda, A., Chládová, Z., and Pišoft, P. Comparative validation of statistical and dynamical downscaling models on a dense grid in central europe: temperature. *Theoretical and Applied Climatology*, 120(3):533–553, 2015.

Hyun, S. and Heo, J.-P. Varsr: Variational super-resolution network for very low resolution images. In *European Conference on Computer Vision*, pp. 431–447. Springer, 2020.

Jebeile, J., Lam, V., and Räz, T. Understanding climate change with statistical downscaling and machine learning. *Synthese*, 199(1):1877–1897, 2021.

Jo, Y., Yang, S., and Kim, S. J. Srflow-da: Super-resolution using normalizing flow with deep convolutional block. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 364–372, 2021.

Kareem, S., Hamad, Z. J., and Askar, S. An evaluation of cnn and ann in prediction weather forecasting: A review. *Sustainable Engineering and Innovation*, 3(2):148–159, 2021.

Kettle, H. and Thompson, R. Statistical downscaling in european mountains: verification of reconstructed air temperature. *Climate Research*, 26(2):97–112, 2004.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.

Leinonen, J., Nerini, D., and Berne, A. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

Lugmayr, A., Danelljan, M., Gool, L. V., and Timofte, R. Srflow: Learning the super-resolution space with normalizing flow. In *European conference on computer vision*, pp. 715–732. Springer, 2020.

Misra, S., Sarkar, S., and Mitra, P. Statistical downscaling of precipitation using long short-term memory recurrent neural networks. *Theoretical and applied climatology*, 134(3):1179–1196, 2018.

Nicholls, R. J. and Cazenave, A. Sea-level rise and its impact on coastal zones. *science*, 328(5985): 1517–1520, 2010.

Pierce, D. W., Cayan, D. R., and Thrasher, B. L. Statistical downscaling using localized constructed analogs (loca). *Journal of Hydrometeorology*, 15(6):2558–2585, 2014.

Rakotonirina, N. C. and Rasoanaivo, A. Esrgan+: Further improving enhanced super-resolution generative adversarial network. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3637–3641. IEEE, 2020.

Rasmussen, R. and Liu, C. High resolution wrf simulations of the current and future climate of north america, 2017. URL `https://doi.org/10.5065/D6V40SXP`.

Sekiyama, T. T. Statistical downscaling of temperature distributions from the synoptic scale to the mesoscale using deep convolutional neural networks. *arXiv preprint arXiv:2007.10839*, 2020.

Sun, L. and Lan, Y. Statistical downscaling of daily temperature and precipitation over china using deep learning neural models: Localization and comparison with other methods. *International Journal of Climatology*, 41(2):1128–1147, 2021.

Vaughan, A., Lane, N. D., and Herzog, M. Multivariate climate downscaling with latent neural processes.

Vaughan, A., Tebbutt, W., Hosking, J. S., and Turner, R. E. Convolutional conditional neural processes for local climate downscaling. *Geoscientific Model Development*, 15(1):251–268, 2022.

Villén-Peréz, S., Heikkinen, J., Salemaa, M., and Mäkipää, R. Global warming will affect the maximum potential abundance of boreal plant species. *Ecography*, 43(6):801–811, 2020.

Wang, L., Chen, J., and Wu, L. Impact of winter sst anomaly in tropical eastern pacific on following summer tropical cyclone precipitation in china. *International Journal of Climatology*, 40(2): 739–749, 2020.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.

Widmann, M., Bedia, J., Gutiérrez, J. M., Bosshard, T., Hertig, E., Maraun, D., Casado, M. J., Ramos, P., Cardoso, R. M., Soares, P. M., et al. Validation of spatial variability in downscaling results from the value perfect predictor experiment. *International Journal of Climatology*, 39(9):3819–3845, 2019.

Wilks, D. S. Stochastic weather generators for climate-change downscaling, part ii: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 3 (3):267–278, 2012.

A Appendix

A.1 Multi-scale Climate Dataset for the Continental United States

In order to better facilitate future work in climate downscaling, we have aggregated and preprocessed precipitation and temperature data at three different resolutions. This combination provides training data for three possible downscaling tasks, as shown in table 6. Because of limited memory and preprocessing time, we chose to keep the maximum WRF resolution capped at $\frac{1}{16} \times \frac{1}{16}$ degrees, as well as restrict climate variables to only accumulated precipitation (mm) and maximum temperature (K) (for reference, the size of raw WRF temperature data before downsampling was roughly 200 GB).

Table 6: Downscaling options provided by our current dataset

Task	Zoom	Overlap	Variables	#Samples
ERA–>CPC	3x	1/1/00 - 12/31/18	precip.	6,935
CPC–>WRF	4x	10/1/00-12/31	precip.	2,282
ERA–>WRF	12x	10/1/00-12/31	precip., temp.	2,282

A.1.1 Preprocessing

While each of our three data sets provides data for the continental United States, their coordinates do not align exactly (ERA and CPC have a constant offset, but WRF has a non-regular grid). Therefore, before training, we perform nearest neighbor interpolation to a regular grid closely matching (no higher resolution than) the highest resolution dataset. Additionally, we downsample the hourly data from ERA and WRF datasets to daily measurements by taking summations and maxima for precipitation and temperature data, respectively.

To increase the temporal overlap between the CPC and ERA-iterim datasets, we separately preprocess a real-time version of the CPC dataset from 2007-2018[2] and combine it with the original.

Another attempt at increasing the number of samples is through data augmentation. We take 5 different 10×10 degree patches of the United States: Northwestern, Northeastern, Midwestern, Southwestern, and Southeastern (see Figure 1). Completely randomizing the training patches and allowing for overlap may allow for a much larger multiplier on the base dataset size, at the risk of overfitting. While rotations of the data were also considered, the transformation may interfere with the network's coordinate-based conditioning (rotations would result in pixels being mislabeled as the wrong coordinate) and so it was not performed.

To generate the coordinate grids used in interpolation, we divide the five selected 10×10 degree regions into a regular grid according to the desired highest resolution. We generate three differently sized image sets this way: 40×40 pixel images at 0.25 degree resolution, 100×100 pixel images at 0.10 degree resolution, and 160×160 pixel images at $\frac{1}{16}$ degree resolution. The details of these image collections are provided in table 7.

Table 7: Collections of climate data at different image sizes

Collection	Datasets Incl.	Size
precip-40	ERA, CPC	268 MB
precip-100	ERA, CPC, WRF	1.81 GB
precip-160	ERA, CPC, WRF	4.66 GB
temp-160	ERA, WRF	3.40 GB

A.1.2 Feature Analysis

We justify our data alignment by Figure 7, where we see that our spatially averaged ERA-iterim, CPC, and WRF datasets align closely. In Figure 9, we see that our higher resolution data have larger extremes than the coarser datasets (as expected).

[2] `https://psl.noaa.gov/data/gridded/data.unified.daily.conus.rt.html`

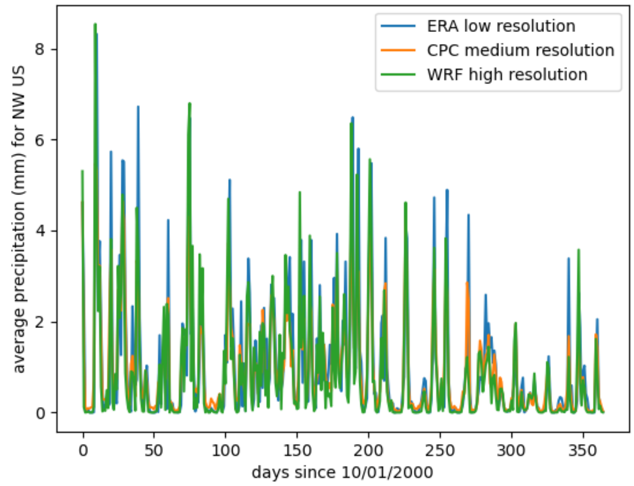


Figure 7: Average (spatial) precipitation over a large region of the northwest during a period of 1 year, where high, medium and low-resolution datasets match closely, with the low-resolution data consistently biased higher during periods of heavy rainfall.

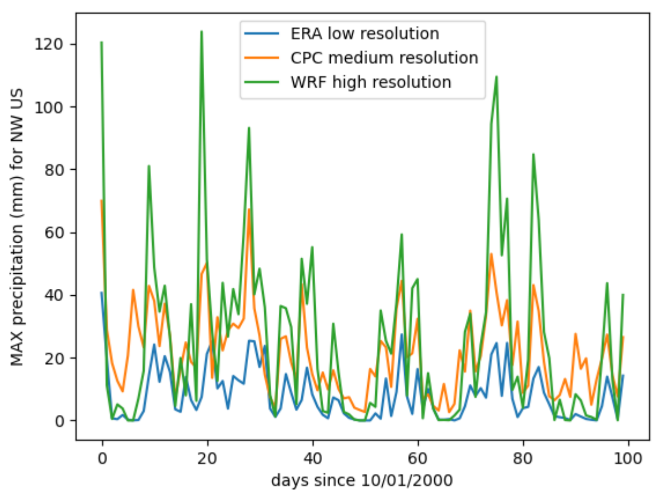


Figure 8: Maximum (spatial) precipitation over a large region of the northwest during a 100-day period. Higher resolution data consistently has larger maxima than coarser data.

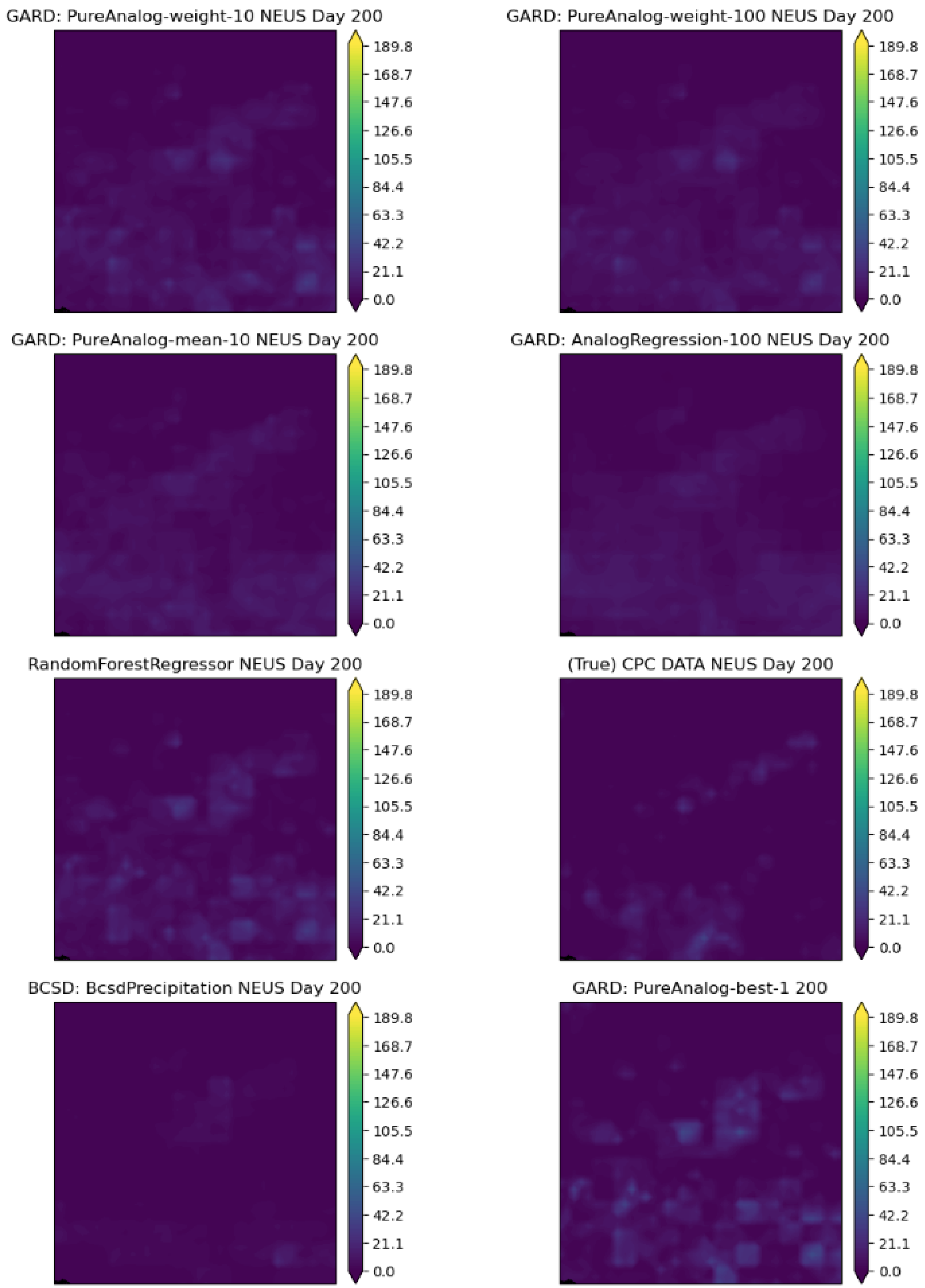


Figure 9: The predicted precipitation heat map for the year 2001 was trained on the year 2000 at 40x40 resolution. It also includes the true high-resolution precipitation data for the year 2001. The data is for the northeast US.