Dear Editor and Reviewers,

We greatly appreciate your handling of our manuscript, *Fast and accurate learned multiresolution dynamical downscaling for precipitation*. We especially thank the two anonymous referees for their detailed and constructive comments and suggestions, which helped improving the manuscript significantly. To respond to reviewers' comments and suggestions, we mark all the changes in purple in our revised manuscript. Major efforts include that, we made quantile-quantile (q-q) plots; we made J-S distance maps, we updated Figs 7-9 with relative error maps; we conducted another experiment to increase the number of parameters in Direct models for investigating whether the Encoded models' improvement is due to model size or model architecture; we have also cleaned up our source code and added more comments. In addition, we updated Figure 3 and the numbers in Table 2. Please find our one-on-one responses below to the comments and suggestions provided by the reviewers.

Sincerely,

Jiali Wang, Zhengchun Liu, Ian Foster, Won Chang, Rajkumar Kettimuthu, Rao Kotamarthi.

1 Comments and Response to Reviewer #1

Comments: MSE: I am not quite sure why the authors define MSE as $\frac{1}{N}\sum_{i=1}^{N}\left(Y_i-\overline{Y}\right)^2$, where N is the total grids. A more sensible MSE would be $\frac{1}{N}\sum_{i=1}^{N}(Y_i-Y_{i,GroundTruth})$. Also, I am not sure there is a need to calculate MSE at each timestep unless the authors plan to explore how MSE varies with time. Therefore I would suggest the authors to calculate MSE as $\frac{1}{N\times T}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{i,t}-Y_{i,t,GroundTruth})^2$ where T is the total time steps during the testing period.

Response: Thanks for pointing this out. We have revised the equation for MSE (eq. 2) in Page 11, as suggested. We computed MSE for each time-step, because we assess the performance of our deep-learning (DL) models in different percentiles of the precipitation. For example, in Table 2, we present the MSE for the 50th and 99th percentiles of precipitation. We agree that, if the variation of MSE with time is not of an interest, then the MSE could be calculated for all grid cells at all times. Here is what we have in the revised manuscript:

$$\ell_{mse}=\frac{1}{N}\sum_{i=1}^{N}\left(Y_i^H-G_\theta\left(Y_i^L\right)\right)^2, \tag{1}$$

where N is the total number of grid cells over the study domain, Y_i^H and Y_i^L are the precipitation at grid cell (i) simulated by WRF at high (12 km) and low (50 km) resolution respectively. Y_i^H is used as Ground Truth in this case. G_θ is the deep neural network parameterized with θ that models the difference between low and high resolution simulations.

Comments: Precipitation Distribution: I don't think J-S distance used here is very informative, a single number does not tell us how two distributions are different (and such assessment can be made, at least qualitatively, using log-pdfs shown in Fig. 6). However, it would be of interest to

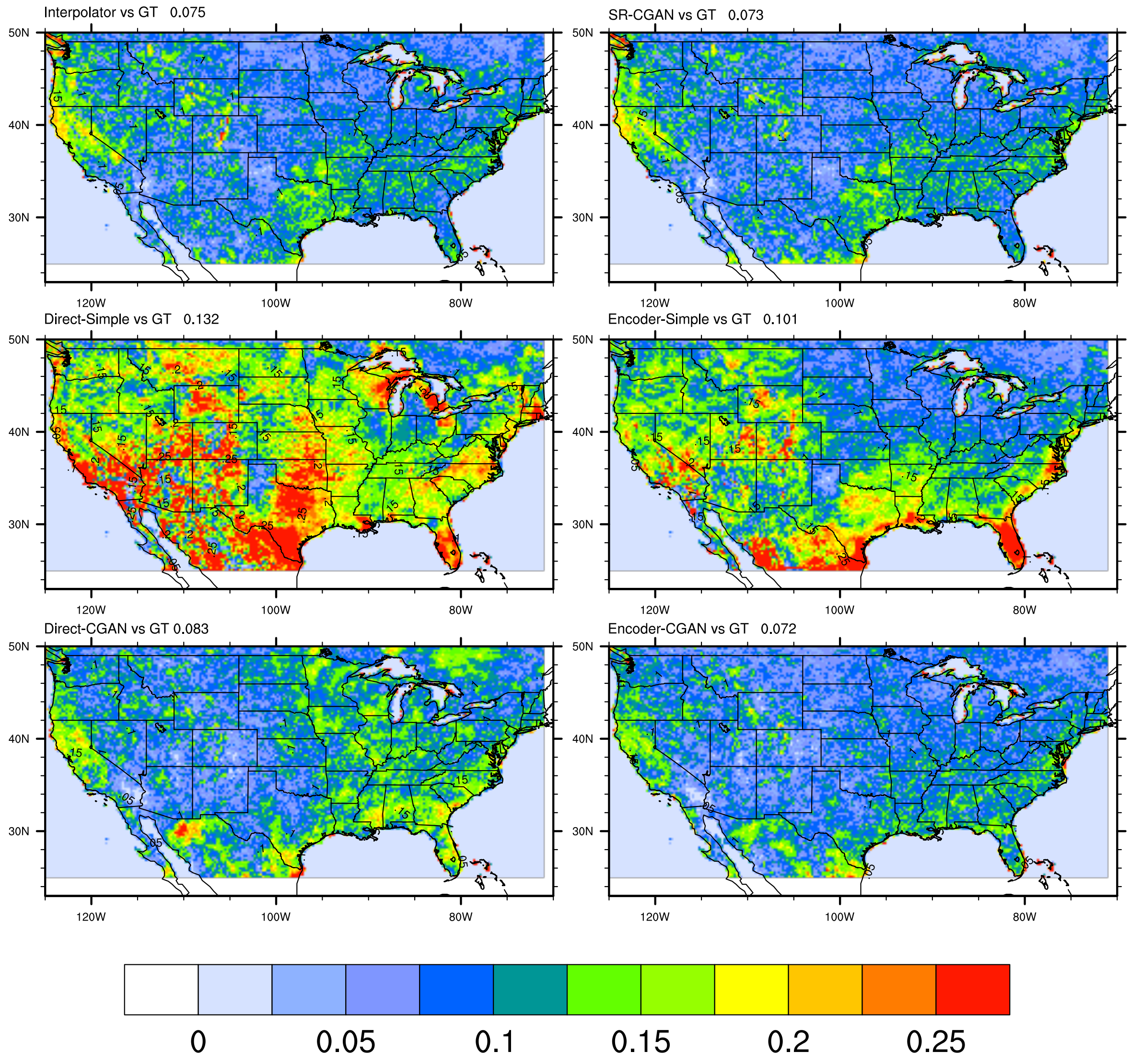


Figure R1: J-S distance measuring the similarity of the PDFs between Ground Truth and six predictive models for the testing period (October–December).

show spatial maps of J-S distance at each grid cell across different methods as it could potentially provide additional information for sub-region assessment in terms of fitted distributions. I would also suggest to replace the log-pdf curves in Fig. 6 by QQplots, that is, for each region, plot true (empirical) quantiles against the predicted quantiles. Quantile values are more directly interpretable than these log-density curves.

Response: According to reviewer's suggestions, (1) we have added the J-S distance maps which show the similarity of the PDFs of each grid cell between Ground Truth and the six different downscaling approaches. As shown in Figure R1, compare with Interpolator and SR-CGAN, while the Simple models produce the largest distances from the Ground Truth, the Encoded-CGAN produces the smallest distances over the entire CONUS, indicating that Encoded-CGAN can generate closer precipiation distributions to Ground Truth than the other five downscaling approaches.

(2) We have made Q-Q plots using all the non-zero precipitation data and the precipitation data greater than 8mm/3hr over each of the seven subregions. We found that, the Q-Q plot using all non-zero precipitation is dominated by the majority of the data which are not really at the tails; while the Q-Q plot using precipitation greater than 8mm/3hr can capture the information at the tail, and we also found them showing a similar conclusion as the PDF plots (see Figure R2). That is, the CGAN models usually can capture the very extremes better than the other downscaling approaches. Since we are interested in evaluating the tails of the distributions predicted by different DL models, the PDF plots are more suitable than Q-Q plot. For example, if there are two distributions with similar tails, but they are different in other parts of the distribution, then Q-Q plot may conclude that these two distributions are very different (see Figure R3). Therefore, we kept our PDF plots and the J-S distances (Table 3) that are used to quantify the similarity in PDFs generated by different DL models.

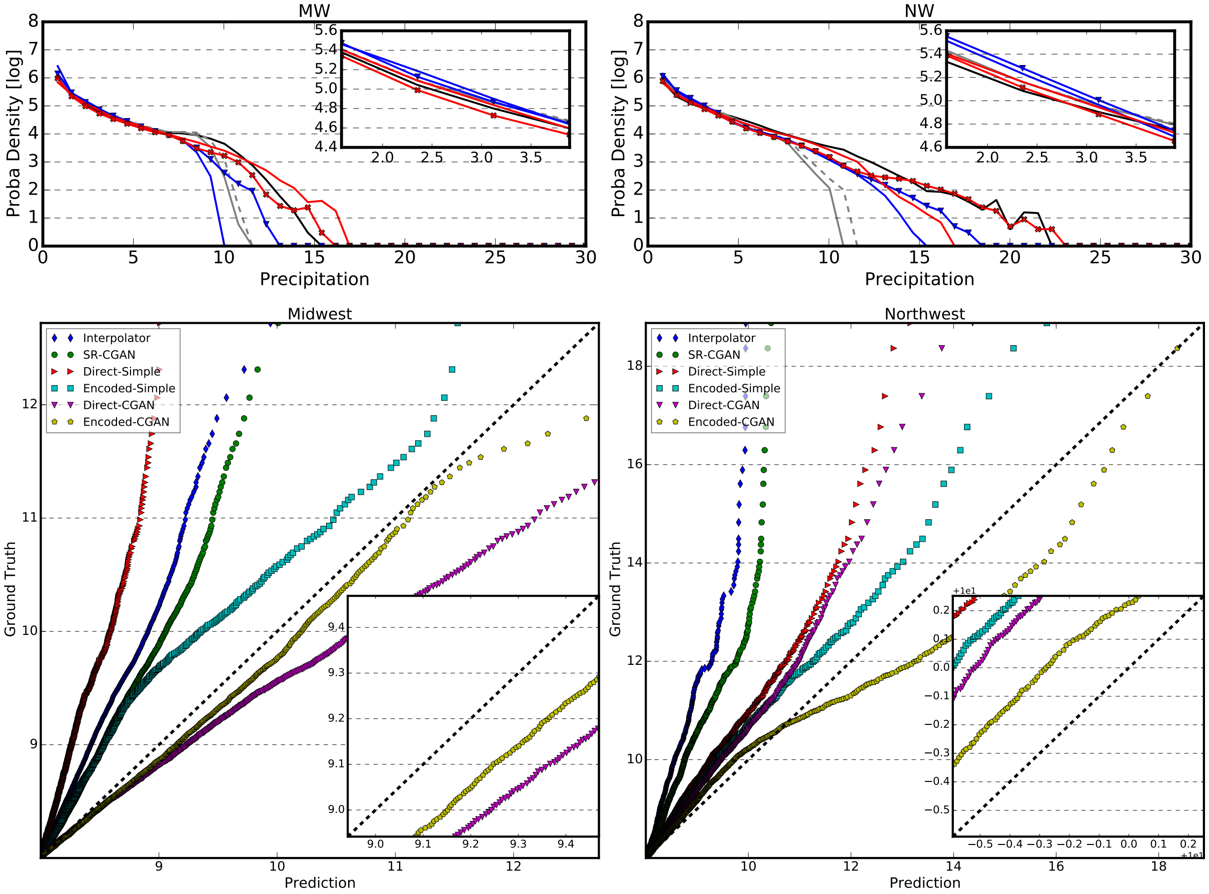


Figure R2: PDF (top) and Q-Q (bottom) plots over Midwest and Northwest. In PDF plots, the red line with cross is Encoded-CGAN, and the blue line with triangles is Encoded-Simple. Both PDF and Q-Q plots indicate that CGAN models are closer to Ground Truth over Midwest, and Encoded models are closer to Ground Truth over Northwest.

Comments: * page 2, line 42-43 “Running an RCM is computationally expensive, however, and typically cannot be applied to large ESM ensembles”: Please include reference for the Canadian Re-

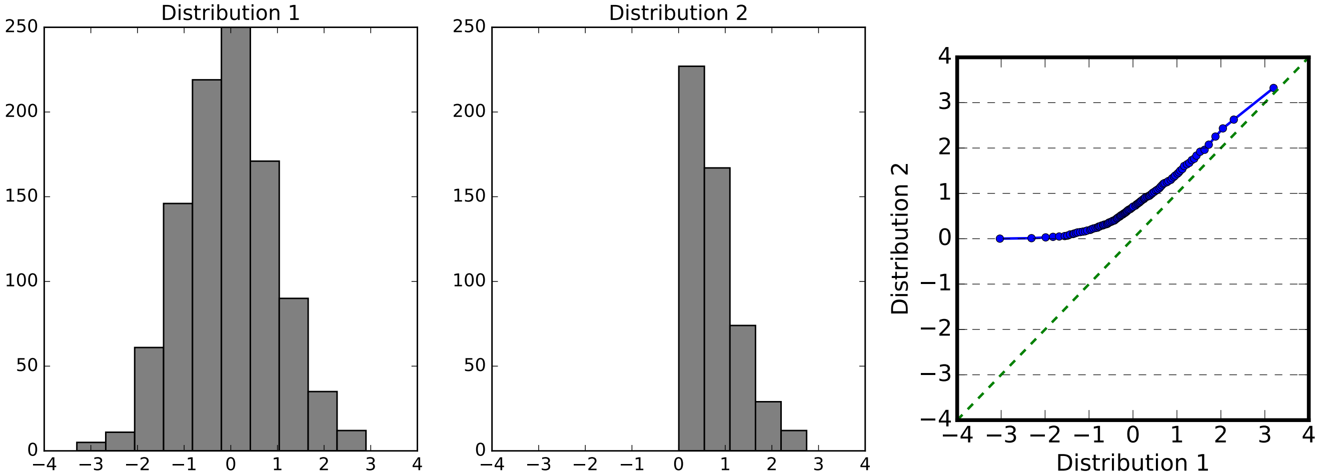


Figure R3: An example of two distributions plotted in histogram and Q-Q plots. These two distributions are similar on the right side (greater than 0) but is different on the left side.

gional Climate Model Large Ensemble here, for example, Kirchmeier-Young, M. C., N. P. Gillett, F. W. Zwiers, A. J. Cannon, F. S. Anslow, 2018: Influence of human-induced climate change on British Columbias extreme 2017 fire season. Earth's Future, 7, 2-10. https://doi.org/10.1029/2018EF001050.

Response: Thanks and the citation is included.

Comments: * page 4, line 100 "The data used in this study are one-year outputs...": Did the authors apply their methods to another one-year outputs to check if (qualitatively) similar conclusions can be obtained?

Response: The neural networks we developed are only based on one year of data which is not able to capture the inter-annual variability. Thus when the trained model is applied to a different year, we expect the model performance could be unsatisfied. On the other hand, if we conduct both training and testing for a different year (e.g., use first 9 months for training data, and the rest for testing), we expect the conclusion will be similar to what we got here. That is, our multi-resolution model can produce high-resolution modeled precipitation data with comparable statistical properties but at greatly reduced computational cost. This training and verification for a new year will take significant amount of effort as we did before with the data from year 2005. It is possible though, that we may end up with a slightly different set of hyperparameters that achieves the best results. This will require careful tuning and testing of the model development. With more computational resource available in near future, we would like to train our neural networks with multiple years of data to consider the inter-annual variability; we also would like to apply our techniques to other coarse resolution models, such as the state-of-the-art earth system models that are on grid spacing of 50-100 km. We added this discussion in the section of Summary and Discussion.

Comments: * page 13, Table 2: These information can be well summarized by a scatter plot by putting regions on x-axis and MSEs on y-axis with different color/line symbol combinations for these CNN models.

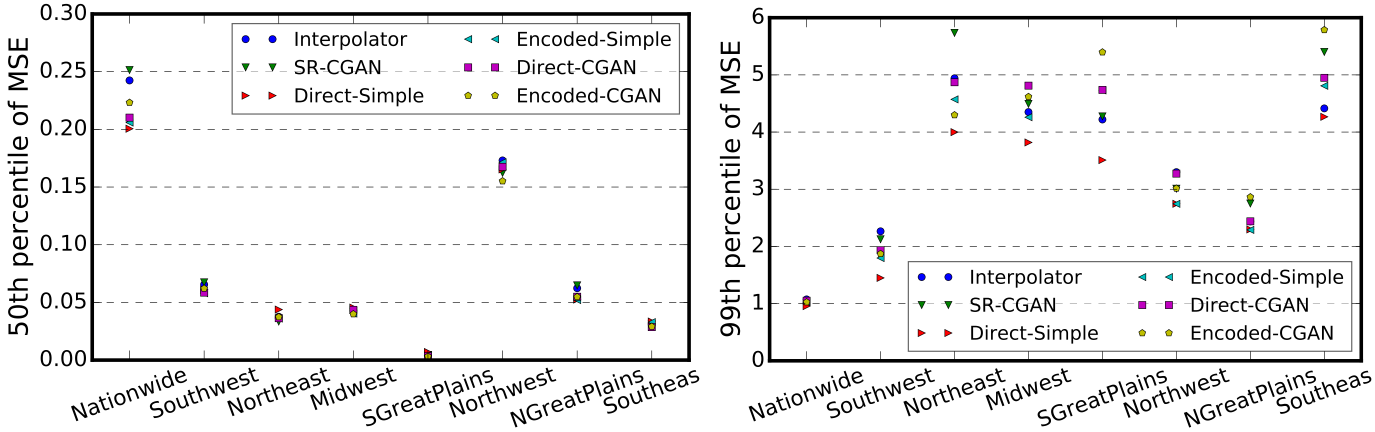


Figure R4: MSEs, calculated across all grid cells over the entire CONUS and seven subregions, at the 50th and 99th percentiles picked from all timesteps.

Response: Thanks for the suggestion. We have tried the scatter plots, as shown by Figure R4. We can see that, because the MSEs are larger over Northwest and CONUS than those over the other six subregions, putting them in one plot make it difficult to read the differences between different downscaling approaches. We thus kept the original table, but we updated the numbers in Table 2 as we found some numbers are incorrect. In addition, we would like to emphasize that, in Table 2 we can see that MSE does not seem to be a sufficient metric for evaluating precipitation, especially its spatial patterns, because MSE tends to evaluate the spatial average instead of the small-scale features (e.g., heavy precipitation), which is more important in real-situation applications such as risk assessment of heavy precipitation or flooding.

Comments: Figs. 7-9: I would suggest the authors to plot relative error maps to better compare between different CNN model fits.

Response: Thanks for the suggestion. We have updated these figures (Figs. 8-10 in revised manuscript) with relative error maps. An example of the difference in top 5% precipitation between Ground Truth and all six downscaling approaches is shown by Figure R5. We can see that both Interpolator and SR-CGAN underestimate the high precipitation over Northwest, but overestimate the precipitation over central and eastern US, indicating smoother spatial patterns than those generated by the four DL models we developed. Both the Simple models and the CGAN models reduce the bias (e.g., underestimation shown in Interpolator and SR-CGAN) over Northwest and over south central and Southeast (e.g., overestimation shown in Interpolator and SR-CGAN). We see similar improvements for standard deviation and monthly averaged precipitation.

Comments: * Fig 10: I would recommend to use QQplots here for easier comparison.

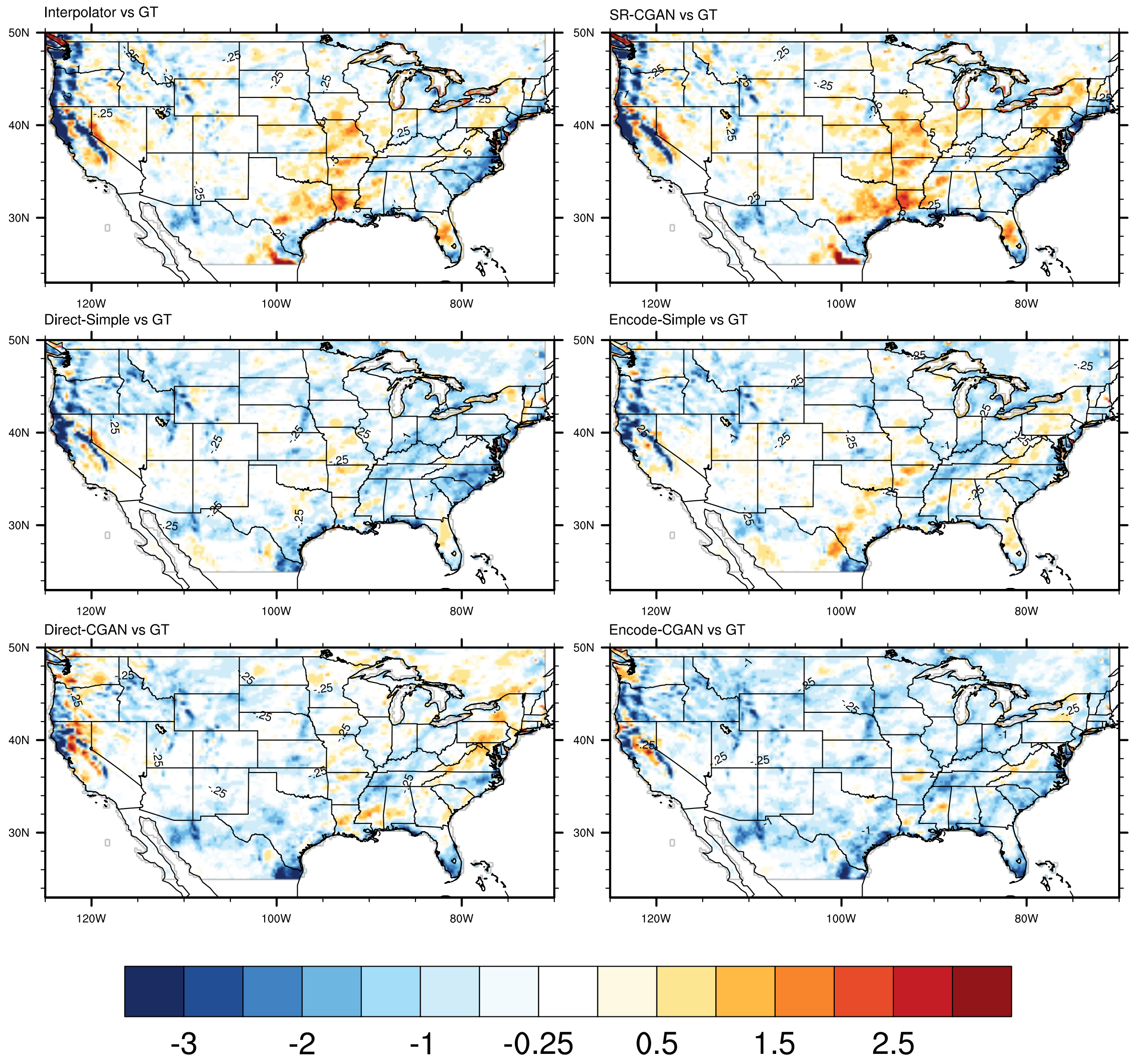


Figure R5: Differences in top 5% (averaged across the 95th percentile to maximum) of the precipitation amount (mm/3 hr) during the testing period (October–December).

Response: As we responded to the previous comment about Q-Q plots, because the distributions of the precipitation features (intensity, size, duration, and frequency) from each DL model are not exactly the same, and we are interested more in the extremes, we kept the histogram plots for each of the characteristic for the rainstorm. Please note here we only tracked the events that have precipitation amount greater than 10mm/3hr, because we focus on the very extremes due to their large impacts. It is the same reasons that we use PDF plots for model evaluations for the extremes.

2 Comments and Response to Reviewer #2

Comments: How is scaling/pre-processing/post-processing of the data handled? It looks, based on the code, like the output activation from the CNN is a sigmoid (logistic) function, so how are the precipitation data mapped to/from the (0,1) range? Do you take any steps to account for the very skewed distribution of precipitation? Does the sigmoid output cap the maximum possible output precipitation? If there is a non-linear transform used to map from the CNN outputs back to dimensional precipitation values how does this affect using an MSE loss function? These considerations should be addressed in the manuscript.

Response: Thanks for your questions. The sigmoid function you saw in the source code is for the spatial and channel attention module, i.e., the CBAM as illustrated in Fig. 3. The activation for the output layer of the generator is linear. The sigmoid function is not for precipitation data, and we did not map precipitation to the (0,1) range. To avoid confusion, we added comments in the source code to clarify the coding structure at the first place. We have also renamed several functions to make them more readable. Please find the refined source code available in the same repository `https://github.com/lzhengchun/DSGAN`.

Comments: Lines 138, 166 and Fig 3: It is not clear how the elevation data are passed to the CNN. Line 138 says it is concatenated to the low-res inputs and Line 166+Fig 3 make it seem that it is only concatenated to features derived from the other fields near the end of the network.

Comments: Line 138: *A friendly suggestion regarding the elevation data* If you are only concatenating the DEM data near the end of the network like Fig 3 seems to suggest, you may want to consider adding the elevation data earlier in the model. It appears that after the DEM data are concatenated they are only passed through one layer alongside the features learned from the other inputs before the model output. I suspect this could limit the CNN's ability to learn the complicated non-linear relationships between the DEM and the other fields needed to estimate orographic precipitation (letting the information about orography flow through the channel attention blocks might be particularly helpful). I would recommend finding a way to add the DEM data earlier in the network, perhaps you could use convolutional downsampling or a pixel shuffle instead of just 2d-avg to get it down to the same resolution as the other inputs without losing much information. I think the paper is fine without doing this of course, since you have already achieved very good performance.

Response: Thanks for your comment and suggestion. Here we are responding them together since they are referring the same figure (Figure 3) about the elevation or DEM data. We realize that our Figure 3 did not fully represent our actual implementation – thanks for pointing it out! As you may have seen in our open source code, the elevation data is concatenated to features after two upsampling operations when the width and height of features match the elevation data. After the concatenating the elevation and the other variables (T2, IWV and SLP) at 12 km, we actually have four inception boxes before the output layer. We have updated Figure 3 in the revised manuscript. It is also shown by Figure R6.

We have also corrected Line 138 by deleting the 'elevation'. It reads now "we directly stack all

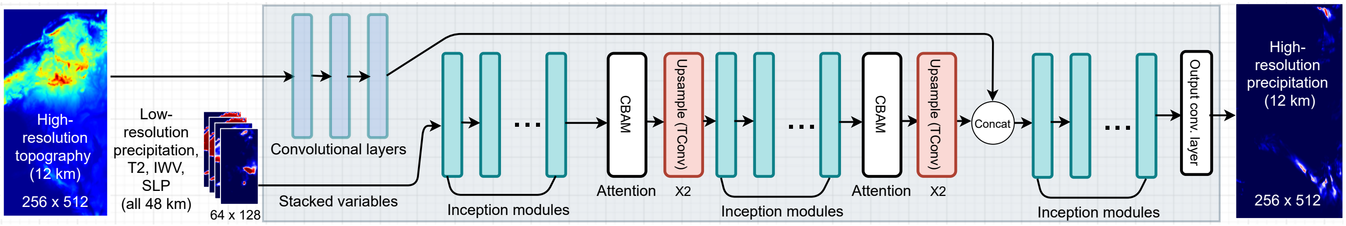


Figure R6: Model architecture for Direct-Simple and the generator of Direct-CGAN. CBAM=Convolutional Block Attention Module. TConv=Transposed convolution.

selected variables (precipitation, T2, IWV, SLP) to form a three-dimensional tensor as input to the CNN model;"

Comments: Eqs 1 and 2: Please explain what 'D' is other than just the discriminator output. Your code looks like it is binary cross entropy applied to a sigmoid output with binary class labels. This info should be included here.
Comments: S 2.5: No information about the discriminator network architecture is given.

Response: We are responding these two comments together since they are both about the discriminator. Let `CxKySz` denote a convolution layer with `x` channels, kernel size of `y` and striding of `z`. Our discriminator network architecture is `C64K4S2-C128K4S2-C256K4S2-C512K4S2-C512K4S2-C1K4S2`. Batch Normalization are applied for outputs of layer 2 to layer 5. We used leaky ReLU activation function with a negative slope of 0.2 for all layers except the last layer (the output layer) for which the sigmoid activation is used. In the revised manuscript, we added the suggested sentence here: "*D* is the discriminator, and is binary cross entropy applied to a sigmoid output with binary class labels."

Comments: Eq 3. Does not seem like an appropriate error metric for this problem. Why is the difference between the SR/downscaling and a spatial mean of precipitation used? Shouldn't this be something like: MSE = (1/N) $\sum_{i=1}^{N}(Y_i^{CNN} - Y_i^{GT})$? (GT = ground truth). I believe that is the type of MSE the CNN is optimizing. To me, Eq 3 looks more like the variance of the CNN output computed about the ground truth mean rather than the MSE. Also, just practically speaking, I don't think adequately estimating total precipitation averaged over all of conus really matters if we can't accurately say where within CONUS it is happening.

Response: Thanks for pointing this out. We have corrected eq. 3 as suggested by both reviewers. We also agree that, it is very important that a model can predict the precipitation over specific regions or locations. To evaluate this regard, we conducted model evaluation over each of the subregions, and also over each grid cell by looking at geospatial patterns across the entire CONUS in Section 3.3 Geospatial analysis of other measures, illustrated by Figures 7-10.

Comments: S 3.2: What are the relative numbers of trainable parameters between the models?

Do the encoded CNNs improve performance (in terms of the PDFs at least) because of the model architecture difference or simply because they have more convolutional kernels?

Response: The number of parameters for the encoded-CGAN (1,487,683 trainable parameters) is a little (2.74%) more than the direct-CGAN (1,447,971 trainable parameters) as encoded-CNN has the encoder layers shown in Fig. 5. In order to investigate whether the improvement is due to the larger number of parameters or the neural network architecture, we conducted an experiment by adding more inception boxes (the encoder also uses the inception box) to the direct-CGAN to make the parameters similar as the encoded-CGAN. It ended up with 1,499,363 parameters, a little more than the encoded-CGAN. Let's name the new model as Direct-Big-CGAN. The median MSE of all grid cells (as shown in 1st column of Table 2) are almost identical between the Direct-Big-CGAN and the original direct-CGAN (0.21043 vs. 0.21015).This result indicates that even with a larger number of parameters, the performance of the direct-CGAN is still the same. Therefore, the improvements of encoded-CGAN compared with direct-CGAN over many subregions for PDFs (shown in J-S distance maps) are most likely due to the neural network architecture not the larger number of parameters. We added this discussion to the section of Summary and Discussion.

Comments: Title: what is meant by "multi-resolution"? doesn't this only produce 12km outputs?

Response: Our output from the deep learning based downscaling approach only produce 12km outputs for this particular study, you are correct. However, by "multi-resolution", we mean that our approach is developed based on datasets that are on more than one spatial resolution. This is contrast to the state-of-the-art Super-Resolution technique, which upsample a high resolution data, and develop the SR model using the upsampled low and high resolution data which are based on the same data source. The strength of our approach is that during the training, the neural network can model the difference between two completely different datasets that are generated by low and high resolution numerical simulations. This is particularly useful for climate model or earth system model output, because they can generate different output with different spatial resolutions. As we explained in the 2nd paragraph of Section 2.1, the difference between two simulations that are run at different spatial resolutions involve terrain effect, effects of time steps, impacts of slightly different model domain coverage. Therefore, modeling the difference between these two datasets is more challenging but can potentially produce better results when apply it to a low resolution model simulation. Note that this low resolution data was used during the training.

Comments: General: There's inconsistent terminology used in the ML and atmospheric science fields: "upsampling" and "down-scaling" both refer to resolution increases while "down-sampling" and "up-scaling" both refer to resolution decreases. It would be helpful to make this clear somewhere in the manuscript to prevent confusion for a reader who is only familiar with one of the fields.

Response: Thanks for the comment and agreed. We added a clarification to the end of Introduction. It reads "It is noteworthy that "upsampling" and "downscaling" both refer to resolution increases (from low to high resolution), while "downsampling" and "upscaling" both refer to resolution decreases (from high to low resolution). Since this is an interdisciplinary study, different terms are used in different contexts but refer to the same meanings”

Comments: Abstract/Line 73: When reading the abstract, it’s certainly implied but I don’t think it is explicitly stated that a key difference between your method and the CNN-SR method is that one uses high- and low-resolution simulations to train and the other uses a high-res simulation with statistical up-scaling to train.

Response: Yes, that is the key difference, and that is also what we mean by “multi-resolution” in the title. We train DNNs to capture the difference between low- and high-resolution simulations. Their difference is not only in spatial resolution but also in geospatial patterns, as shown in Figure 2; CNN-SR, on the other hand, upscales a high-resolution image to low-resolution and then generate a high-resolution data that is close enough to the original high-resolution data. So the only difference in CNN-SR’s training data is the spatial resolution. In the revised manuscript, we emphasize this regard in both Abstract and Line 73. Abstract reads now: “The key idea is to use combination of low- and high- resolution simulations (differ not only in spatial resolution but also in geospatial patterns) to train a neural network to map from the former to the latter.”. Line 73 reads now: “Two datasets (i.e., low- and high-resolution simulations) rather than one (i.e., only high-resolution with its upscaling) are used to develop the new CNN-based downscaling approach.”

Comments: Line 93: “precipitation images” – is an odd term since the data are really only similar to an image in that they are gridded/raster data. Presumably the CNN is operating on the actual precipitation data and not the RGB images of it.

Response: We agree with the reviewer and have changed the “precipitation images” to “precipitation visualization” as we basically compare it visually there. We also changed “precipitation images” in other contexts to “precipitation” as we do not really talk about image or visulization but just the precipitation data itself.

Comments: Line 140: again I don’t think referring to the data as images is helpful since they are not images. Something like: “This approach of stacking the variables as different input channels has been used in other downscaling studies. . . ” would be fine.

Response: Corrected.

Comments: S 2.3: Did you provide any source of random variability to the CGAN?

Response: By “random variability”, do you mean the latent variable input to the generator? We do not provide this because as we mentioned in the manuscript, ”..we use actual precipitation amounts and the conditional variables as inputs, forming a conditional GAN (CGAN) framework for training the generator.”. The discriminator is a helper here to provide a better loss function for the generator.

Comments: Table 2: Units?

Response: These are MSE for precipitation amount at each time step. We use 3 hourly precipitation data, so the MSE here is in mm/3h. Added in Table 2 caption.

Comments: SS2.5 It feels disorganized to introduce the GAN model in a section labeled "loss functions." I suggest just having a section early on that briefly introduces each model to help the reader keep everything straight. Then proceed to discuss specifics of implementation / loss functions etc.

Response: Thanks for the suggestion. We first describe the model architecture, and they are Direct-Simple and Encoded-Simple models. Then we describe the loss functions we use, and they are MSE and CGAN. Accordingly, we introduce the models Direct-CGAN and Encoded-CGAN. So the first part of our model names is about architecture, and the second part is about loss function.Overall, we feel it flows well, so we kept the original structure.

Comments: Line 10: Awkward sentence

Response: Corrected.

Comments: Line 21: change to something like: "ESMs cannot fully resolve cloud processes." the way it's currently worded makes it sound like the models don't include clouds at all

Response: Revised, and it reads now: "Such resolutions are not sufficient to fully resolve critical physical processes such as clouds.."

Comments: Line 45: one number is formatted with a comma and one without

Response: Corrected.

Comments: Line 57: GAN-based SR does not necessarily improve pixel accuracy over a conventional CNN, they improve feature loss or realism. Maybe just be clear about what you mean by "accuracy"

Response: We added the clarification as suggested.

Comments: 109: “convections” → ”convection”

Response: Corrected.

Comments: F1: Thanks for including this diagram, very helpful

Response: We are glad that the reviewer finds this figure helpful. Once again we would like to emphasize that, the difference between our study and the state-of-the-art Super-Resolution technique is that, we use multi-resolution datasets to train the DL models. Results show that it can decently capture the statistical features when comparing with ‘Ground Truth’ (high-resolution WRF simulation here).

Comments: L 194-195: I don’t think this sentence is correct

Response: Can the reviewer please indicate the problem in this sentence? We are happy to discuss more.

Comments: L 204: “less good” → “worse”

Response: Corrected.

Comments: L 225: wow that’s fast! I’m excited to see algorithms like this used operationally

Response: Thanks and we also find this information encouraging. The speed-up these DL models have achieved provide us opportunities to pursue high spatial resolution simulations as well as large ensemble member simulations to better quantify the uncertainties in the climate models. We sincerely thank the reviewers again for all your constructive comments and suggestions!

Fast and accurate learned multiresolution dynamical downscaling for precipitation

Jiali Wang[1], Zhengchun Liu[2], Ian Foster[2], Won Chang[3], Rajkumar Kettimuthu[2], and V. Rao Kotamarthi[1]

[1]Environmental Science Division, Argonne National Laboratory, Lemont, IL, USA
[2]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL, USA
[3]Division of Statistics and Data Science, University of Cincinnati, Cincinnati, OH, USA

Correspondence: V. Rao Kotamarthi (vrkotamarthi@anl.gov); Zhengchun Liu (zhengchun.liu@anl.gov)

Abstract. This study develops a neural network-based approach for emulating high-resolution modeled precipitation data with comparable statistical properties but at greatly reduced computational cost. The key idea is to use combination of low- and high-resolution simulations (differ not only in spatial resolution but also in geospatial patterns) to train a neural network to map from the former to the latter. Specifically, we define two types of CNNs, one that stacks variables directly and one that encodes each variable before stacking, and we train each CNN type both with a conventional loss function, such as mean square error (MSE), and with a conditional generative adversarial network (CGAN), for a total of four CNN variants. We compare the four new CNN-derived high-resolution precipitation results with precipitation generated from original high resolution simulations, a bilinear interpolater and the state-of-the-art CNN-based super-resolution (SR) technique. Results show that the SR technique produces results similar to those of the bilinear interpolator with smoother spatial and temporal distributions and smaller data variabilities and extremes than the original high resolution simulations. While the new CNNs trained by MSE generate better results over some regions than the interpolator and SR technique do, their predictions are still biased from the original high resolution simulations. The CNNs trained by CGAN generate more realistic and physically reasonable results, better capturing not only data variability in time and space but also extremes such as intense and long-lasting storms. The new proposed CNN-based downscaling approach can downscale precipitation from 50 km to 12 km in 14 min for 30 years once the network is trained (training takes 4 hours using 1 GPU), while the conventional dynamical downscaling would take 1 month using 600 CPU cores to generate simulations at the resolution of 12 km over contiguous United States.

1 Introduction

Earth system models (ESMs) integrate the interactions of atmospheric, land, ocean, ice, and biosphere and generate principal data products used across many disciplines to characterize the likely impacts and uncertainties of climate change (Heavens et al., 2013; Stouffer et al., 2017). The computationally demanding nature of ESMs, however, limits their spatial resolution mostly to between 100 and 300 km. Such resolutions are not sufficient to fully resolve critical physical processes such as clouds, which play a key role in determining the Earth's climate by transporting heat and moisture, reflecting and absorbing radiation, and producing rain. Moreover, ESMs cannot assess stakeholder-relevant local impacts of significant changes in the attributes of these processes (at scales of 1–10 km; Gutowski Jr et al., 2020). Higher-resolution simulations covering the entire globe are emerging (e.g., Miyamoto et al., 2013; Bretherton and Khairoutdinov, 2015; Yashiro et al., 2016), including the U.S. Department of Energy's 3 km Simple Cloud-Resolving E3SM Atmosphere Model (E3SM Project, 2018). They are expected to evolve relatively slowly, however, given the challenges of model tuning and validation as well as data storage at unfamiliar scales (Gutowski Jr et al., 2020).

Downscaling techniques are therefore used to mitigate the low spatial resolution of ESMs. Figure 1 illustrates several approaches. Statistical downscaling is computationally efficient and thus can be used to generate multimodel ensembles that are generally considered to be required for capturing structural and scenario uncertainties in climate modeling (Hawkins and Sutton, 2009; Deser et al., 2012; Mearns et al., 2012; Mezghani et al., 2019). However, statistical downscaling works only if the statistical relationship that is calibrated with the present climate is valid for future climate conditions (Fowler et al., 2007). This "stationarity assumption" cannot always be met in practice (Wang et al., 2018). In addition, typical statistical downscaling is limited by the availability of observations, which may lack both spatial and temporal coverage. Furthermore, observations may contain errors, posing challenges for developing a robust model to project future climate.

Dynamical downscaling, in contrast, uses ESM outputs as boundary conditions for regional climate model (RCM) simulations to produce high-resolution outputs. The RCM typically resolves atmospheric features at a spatial resolution of 10–50 km (depending on factors such as the size of the studied domain) with parameterized physical atmospheric processes that in many cases are similar to those used in the ESMs. This approach has the value of being based on physical processes in the atmosphere (e.g., convective scheme, land surface process, short/long wave radiation) and provides a description of a complete set of variables over a volume of the atmosphere. Running an RCM is computationally expensive, however, and typically cannot be applied to large ESM ensembles (Kirchmeier-Young et al., 2019), especially when simulating at the high spatial resolution required to explicitly resolve the convection that cause precipitation storms. For example, a 30-year simulation over a region covering the central and eastern United States with the Weather Research and Forecasting (WRF V4.1.3) model takes 4, 380 core-hours at 50 km resolution but 374,490 core-hours at 12 km resolution and 5.4 million core-hours at 4 km resolution (i.e., 1,224 times more computing resource than at 50 km) on the Intel Broadwell partition of the Bebop cluster at Argonne National Laboratory.

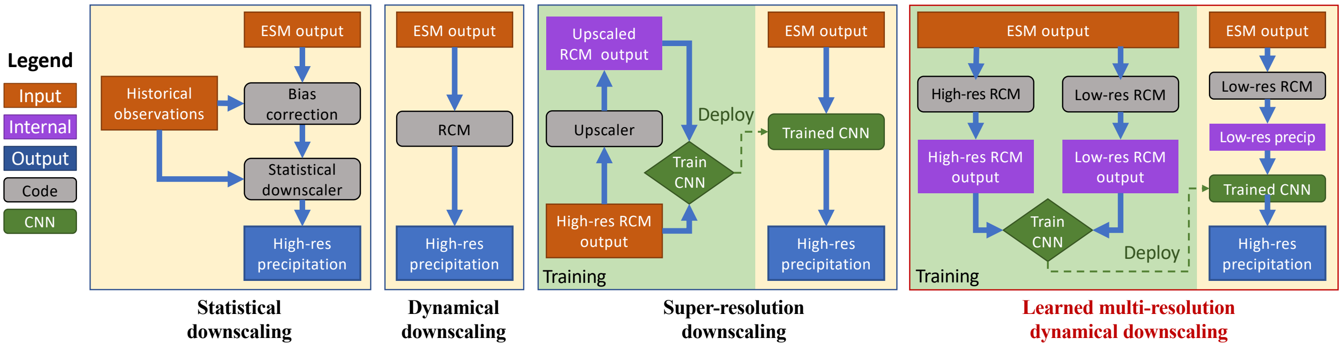
Another recently proposed approach to downscaling uses deep neural networks (DNNs), specifically DNN-based super-resolution (SR) techniques. A DNN consists of several interconnected layers of nonlinear nodes with weights determined by a training process in which the desired output and actual output are repeatedly compared while weights are adjusted (LeCun et al., 2015). DNNs can approximate arbitrary nonlinear functions and are easily adapted to novel problems. They can handle large datasets during training and, once trained, can provide fast predictions (e.g., (Liu et al., 2019, 2020a,b)). In digital image processing, DNN-based SR (Dong et al., 2014; Yang et al., 2014) describes various algorithms that take one or more low-resolution images and generate an estimate of a high-resolution image of the same target (Tian and Ma, 2011), a concept closely related to downscaling in climate modeling. They employ a form of DNN called a convolutional neural network (CNN; LeCun et al., 1998) in which node connections are configured to focus on correlations within neighboring patches. Another DNN variant, the generative adversarial network (GAN; Goodfellow et al., 2014), has been used to improve feature loss or realism of the super-resolution CNNs (Ledig et al., 2017).

SR methods have recently been applied to the challenging problems of downscaling precipitation (Vandal et al., 2017; Geiss and Hardin, 2019) and wind and solar radiation (Stengel et al., 2020), quantities that can vary sharply over spatial scales of 10 km or less depending on location. Downscaling with an SR model proceeds as follows (Vandal et al., 2017; Stengel et al., 2020): (1) take high-resolution data (either climate model output or gridded observations), upscale the data to a low resolution; (2) build an SR model using the original high-resolution and the upscaled low-resolution data; and (3) apply the trained SR model to new low-resolution data, such as from an ESM, to generate new high-resolution data. This general approach has achieved promising results but also has problems. The trained SR model often performs well when applied to a low-resolution data (upscaled from high-resolution data) and compared with the same high-resolution data, capturing both spatial patterns and sharp gradients (Geiss and Hardin, 2019), especially when using a GAN (Stengel et al., 2020). This result is not surprising, given that the high- and low-resolution data used to develop the SR are from the same source. When applied to new low-resolution data such as from an ESM, however, the SR may generate plausible-looking fine features but preserves all biases that exist in the original ESM data, especially biases in spatial distributions and in time series such as diurnal cycles of precipitation, both of which are important for understanding the impacts of heavy precipitation.

Here we describe a new **learned multiresolution dynamical downscaling** approach that seeks to combine the strengths of the dynamical and DNN-based downscaling approaches. Two datasets (i.e., low- and high-resolution simulations) rather than one (i.e., only high-resolution with its upscaling for the SR method) are used to develop the new CNN-based downscaling approach. As shown in Figure 1, these two datasets are both generated by dynamical downscaling with an RCM driven by the same ESM as boundary conditions but at different spatial resolutions. We then use the CNN to approximate the relationship between these two datasets, rather than between the original and upscaled versions of the same data as the SR downscaling does. The output of the CNNs is expected to generate fine-scale features as in the original high-resolution data, because the algorithm is built based on both high and low resolutions. Our goal in developing this approach is to enable generation of fine-resolution data (e.g, 12 km in this study) based on relatively coarse-resolution RCM output (e.g., 50 km in this study) with low computational cost. The combination of high computational efficiency and high-resolution output would allow building more robust datasets for meeting stakeholder needs in infrastructure planning (e.g., energy resource, power system operation) and policy making, where higher spatial resolution (1–10 km) is usually desired. In contrast to statistical downscaling, this approach does not rely on any observations; thus, it can downscale any variables of interest from RCM output for different disciplines. Moreover, because the approach is built on dynamically downscaled simulations, these datasets are not bound by stationarity assumptions. It is noteworthy that "upsampling" and "downscaling" both refer to resolution increases (from low to high resolution), while "downsampling" and "upscaling" both refer to resolution decreases (from high to low resolution). Since this is an interdisciplinary study, different terms are used in different context but refer to the same meanings.

2 Data and Method

This study focuses on precipitation, which is highly variable in time and space and is often the most difficult to compute in ESMs (Legates, 2014). Downscaling of ESMs with RCMs has generally reduced the bias in precipitation projections,



Statistical downscaling

Dynamical downscaling

Super-resolution downscaling

Learned multi-resolution dynamical downscaling

Figure 1. Four downscaling approaches discussed in the text. Orange and blue rectangles are input and output, respectively, of each downscaling approach, and grey rounded rectangles are computational steps. In our learned multiresolution dynamical downscaling approach (right), we use the outputs from low- and high-resolution dynamical downscaling runs driven by the same ESM as boundary conditions to generate {Low, High} training data pairs for a CNN that, once trained, will map from low-resolution dynamically downscaled outputs to a high-resolution.

often due to an increase in the model spatial resolution that allows for resolving critical terrain features such as changes in topography and coastlines (Wang et al., 2015; Zobel et al., 2018; Chang et al., 2020). In addition, precipitation data at high spatial resolution is needed for a variety of climate impact assessments, ranging from flooding risk to agriculture (Maraun et al., 2010; Gutowski Jr et al., 2020). Precipitation data produced by RCMs at each model timestep can be viewed as a two-dimensional matrix or image. However, these precipitation visualization are different from typical photographic images. For example, the precipitation generated by dynamical downscaling at low and high spatial resolutions can be different even if the RCMs used to generate them differ only in spatial resolution. This situation appears often in the precipitation data produced by RCMs running at different spatial resolutions using the same initial and boundary conditions, and it poses a great challenge for developing DNNs for downscaling. In the following subsections we describe in detail the dataset we used for our study, and we discuss our deep learning methods.

2.1 Dataset

The data used in this study are one-year outputs from two RCM simulations using the Weather Research and Forecasting model version 3.3.1, one at 50 km resolution and one at 12 km resolution, both driven by National Centers for Environmental Prediction-U.S. Department of Energy Reanalysis II (NCEP-R2) for the year 2005. These two simulations were conducted separately, not in nested domains, with output every 3 hours for a total of 2920 timesteps in each dataset. Our study domain covers the contiguous United States (CONUS), with 512×256 grid cells for the 12 km simulation and 128×64 grid cells for the 50 km simulation.

The two WRF simulations share the same configuration and physics parameterizations; they differ only in their spatial resolutions. This difference has two direct effects on the precipitation pattern. One is that the higher-resolution results can better resolve physical processes when compared with lower resolution. For example, we can expect the high-resolution simulation to have improved performance for processes that are scale dependent, such as convection and planetary boundary layer physics (Prein and Giangrande, 2020). The other effect is that the higher-resolution model resolves terrain and hence terrain-influenced rainfall, land-sea interface, and coastal rainfall better than the coarse-resolution model does (Komurcu et al., 2018). The difference in spatial resolution also has indirect effects on the precipitation pattern. For example, because these two simulations are not nested, they cover slightly different domains, even though they maintain the same region (CONUS) in the interior. This minor difference in domain position can change the large-scale environment and translate into diverse conditions for the development of the mesoscale processes that produce precipitation (Miguez-Macho et al., 2004). In addition, the difference in spatial resolution leads to the two simulations using different computing timesteps (120 sec versus 40 sec), which can cause precipitation differences due to operator splitting between dynamics and physics in the WRF (Skamarock et al., 2005; Skamarock and Klemp, 2008; Barrett et al., 2019). These factors lead to precipitation differences between the 50 km and 12 km WRF output, as seen in Figure 2, which shows differences between these two datasets in daily January and July means. The 50 km and 12 km data not only have different fine-scale features, such as in the Sierra Nevada and Appalachians, but also have different geolocations of the precipitation, such as that over Texas in July, where the 50 km simulation produces precipitation for 3–5 mm/day but the 12 km simulation produces only 1–2 mm/day. The difference in precipitation between low and high resolution is the biggest challenge for our proposed downscaling approach.

Given these datasets, we need to decide which variables to provide as inputs to our DNN-based downscaling system. Many factors influence the magnitude and variability of precipitation, the focus of this study. Informed by the physics of precipitation, we include, in addition to the low-resolution precipitation and high-resolution topography data used by Vandal et al. (2017) for their SR model, the vertically integrated water vapor (IWV) or precipitable water, sea level pressure (SLP), and 2-meter air temperature (T2) as inputs (Table 1) since we find these variables show high pattern correlations with precipitation along the time dimension. Each variable possesses rich spatial dependencies, much like images, although climate data are more complex than images because of their sparsity, dynamics, and chaotic nature. For each grid cell, we use a fixed threshold (0.05 mm/3 hr) for the minimum precipitation amount, so as to avoid zeros and drizzles being passed to the neural network. Similarly, we define a 99.5th percentile for the maximum precipitation over each grid cell, so as to avoid extremely large precipitation values being passed to the neural network and skews the loss function. Our proposed DNN-based downscaling technique is different from the traditional statistical downscaling methods, particularly regression-based models, which vectorize spatial data and remove the spatial structure.

2.2 Stacked Variables

We design different model architectures and loss functions to make best use of the input variables when training the CNNs to capture the relationship between the precipitation generated by the low- and high-resolution simulations. In this case, we directly stack all selected variables (precipitation, T2, IWV, SLP) to form a three-dimensional tensor as input to the CNN model; see Figure 3. We call the resulting method Direct-Simple hereafter. This approach of directly stacking climate variables as different input channels has been used in other downscaling studies (e.g., Vandal et al., 2017). Unlike those previous studies,

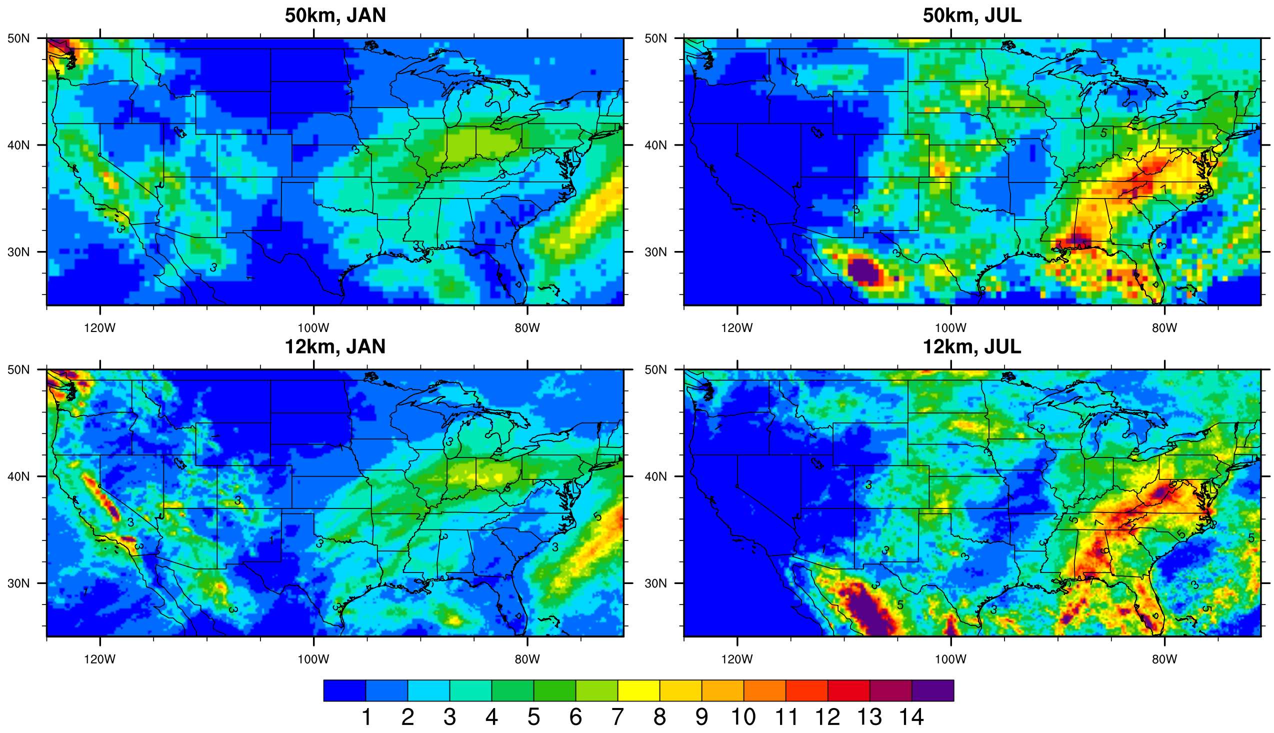


Figure 2. Averaged daily precipitation (mm/day) in January and July of 2005 using 50 km (top) and 12 km (bottom) WRF output. The 50 km precipitation data not only miss fine-scale features, especially over complex terrain such as Sierra Nevada and Appalachians shown by 12 km, but also generate precipitation in different locations, such as July precipitation over Texas.

Table 1. Inputs and outputs for the new CNNs developed in this study. The range column shows the top and bottom 0.1% of the variable over the study domain for the time series of 3-hourly data in year 2005. All data are produced by the WRF model.

Input (units)	Range
50 km, 3-hourly precipitation (mm/3 hr)	[0.05, 13.62]
50 km, 3-hourly SLP (hPa)	[990.97, 1039.34]
50 km, 3-hourly IWV (cm)	[1.56, 116.46]
50 km, 3-hourly T2 (K)	[241.75, 310.35]
12 km, topographic height (m)	[0, 3204.51]
Output	
12 km, 3-hourly precipitation (mm/3 hr)	[0.05, 15.66]

however, we use an inception module as a building block because it can provide different receptive fields at each layer (Szegedy et al., 2015). We use kernels of size 1×1, 3×3, and 5×5 (Figure 4) to build the inception module in order to mitigate the

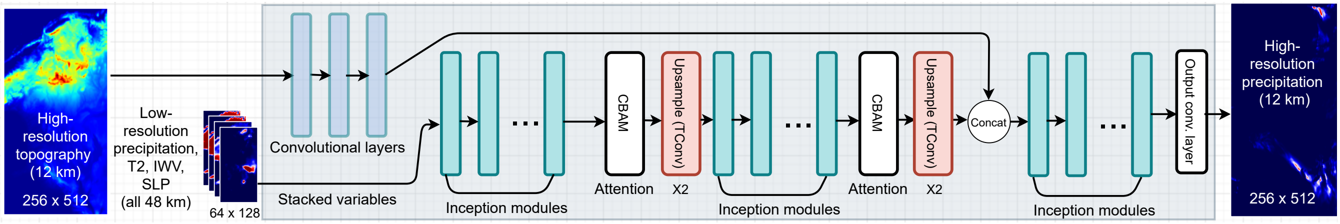


Figure 3. Model architecture for Direct-Simple and the generator of Direct-CGAN. CBAM=Convolutional Block Attention Module (Woo et al., 2018). TConv=Transposed convolution. The inception module is shown in Figure 4.

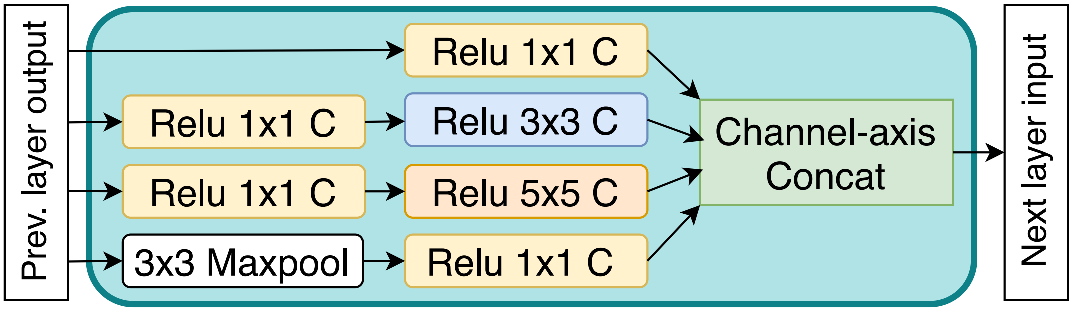


Figure 4. Details of the inception module used in Figure 3.

challenge of learning the relationship between the low- and high-resolution simulations when precipitation occurs in different locations in the two datasets.

From a physical perspective, the inception module makes sense because the precipitation at a location or area is influenced by the conditional variables not only at that particular location but also at adjacent locations depending on the types of weather system. For example, precipitation associated with tropical cyclones (with low SLP centers) over the southeastern United States is usually produced at the eastern or northeastern side of the cyclone center, where the moisture is brought from the Atlantic Ocean or Gulf of Mexico to the northern inland. In addition, stacking the variables considers the coupling effect of all the variables that simultaneously influence the occurrence of precipitation but whose relative importance can be different. Therefore, we apply channel attention (Woo et al., 2018) so that the CNN can learn to focus on the important physical factors that influence the precipitation. On the other hand, the precipitation is extremely sparse in space, with many zeros, posing challenges for the training process. To account for the sparsity of data, we apply a spatial attention (Woo et al., 2018) mechanism to allow the model to learn to emphasize or suppress and refine intermediate features effectively so as to focus on important areas with relatively large precipitation values. This makes physical sense because capturing large precipitation events is critical for climate impact applications, more so than are drizzle or no-precipitation days.

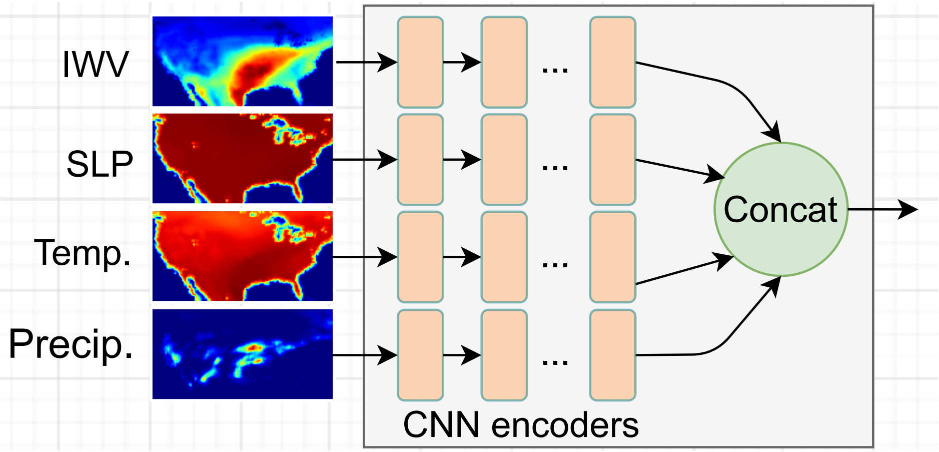


Figure 5. Encoder module used in the Encoded-Simple and the generator of Encoded-CGAN models to prepare the low-resolution input data prior to passing the data to the network of Figure 3.

2.3 Encoded Variables

Stacking all variables as different channels in a CNN is simple and straightforward. However, combining these variables that are significantly different in scales, distribution shapes, sparsity, and units as shown by Table 1 can make the training process challenging. Thus, we develop an encoded variable CNN in which dedicated convolution layers are provided for each variable to extract features before stacking. This ensures that when we stack, the feature maps extracted from each variable have similar characteristics. We call the resulting method Encoded-Simple hereafter and refer to Direct-Simple and Encoded-Simple collectively as the *Simple models*.

Specifically, as shown in Figure 5, we design convolution layers for each of the four variables and stack (i.e., concatenate in the channel axis) their feature maps. We process the topography data similarly but stack the feature maps after the second upsampling operation, that is, when the feature maps size becomes 512×256. From a physics perspective, Encoded-Simple is similar to Direct-Simple except that there are dedicated convolution layers to learn to extract features from each variable during the training process. From the deep learning perspective, this process is friendlier to training, and thus we expect better results from Encoded-Simple than from Direct-Simple. In this study we consider spatial attention for the feature map of each variable before stacking them (i.e., concatenate in the channel axis) and then channel attention (the relative importance of different input variables), because the spatial feature of precipitation is sparse and less uniform and is a critical factor for judging the model performance.

2.4 Super-Resolution Model

To compare the performance of our proposed learned multiresolution dynamical downscaling approach with that of the state-of-the-art SR technique, we develop an SR model based on the original 12 km WRF modeled precipitation and an upscaled

12 km-to-50 km dataset. The SR model development does not need other environmental variables as used in our new CNN approaches. It does not involve the 50 km precipitation data either, thus, the differences between the 12 km and 50 km WRF outputs (as discussed earlier and shown in Figure 2) is not a challenge in the way that it was for the learned multiresolution dynamical downscaling approach. We then (1) apply the trained SR model to the upscaled 50 km dataset and compare the SR-resulted 12 km data with the original 12 km data. This step is to assess the effectiveness of the SR model. For example, when comparing SR-resulted 12 km data and original 12 km data, we find a spatial correlation greater than 0.98 for all the quantiles and almost the same distribution shapes between the two datasets, indicating that the SR model we develop is effective and robust; (2) apply the SR model to our 50 km WRF output for the testing period and compare the resulting SR-generated 12 km data with the original 12 km data. If the SR model can downscale a 50 km dataset to one with similar properties to 12 km WRF output, with much less computational cost than running the 12 km model in conventional dynamical downscaling, then the SR approach is useful for generating high-resolution and high-fidelity precipitation based on low-resolution precipitation

2.5 Loss Functions

A DNN's loss function guides the optimization process used to update weights during training. Thus the choice of loss function is crucial to DNN effectiveness. For the Direct-Simple and Encoded-Simple models introduced above, we first consider two loss functions commonly used in computational vision: the L1 norm (mean absolute error) and L2 norm (mean square error: MSE). Since precipitation data are sparse, those losses may not be able to generate results that are driven primarily by large gradients, such as localized heavy precipitation.

The generative adversarial network (GAN) is a class of machine learning framework in which two neural networks, generator and discriminator, contend with each other to produce a prediction. In our context, the generator network performs CNN by mapping input patches of coarse data to the space of the associated higher-resolution patches. The discriminator attempts to classify proposed patches as real (i.e., coming from the training set) or fake (i.e., coming from the generator network), i.e., basically a binary classifier . The two networks are trained against each other iteratively, and over time the generator produces more realistic fields, while the discriminator becomes better at distinguishing between real and fake data. Therefore, GANs provide a method for inserting physically realistic, small-scale details that could not have been inferred directly from the coarse input images (Liu et al., 2020a).

GANs, as originally formulated, use a vector of random numbers (latent variables) as the only input to the generator. Instead, we use actual precipitation amounts and the conditional variables as inputs, forming a conditional GAN (CGAN) framework (Mirza and Osindero, 2014) for training the generator. The two neural networks (generator and discriminator) are trained simultaneously, with the generator using a weighted average (their weights are hyperparameters) of the ℓ_1-norm (results with the ℓ_2-norm are worse) and adversarial loss as its loss function, defined as

$$\ell(\theta_G) = -\frac{w_a}{m}\sum_{i=1}^{m} D\left(G\left(v_1, v_2, \ldots\right)\right) + \frac{w_c}{m}\left\|Y_i - \overline{Y}_i\right\|_1, \tag{1}$$

where w_a and w_c are weights for the adversarial loss and ℓ_1-norm, respectively; m is the minibatch size; $v_1, v_2, \ldots$ are coarse grained conditional variables including precipitation, IWV, SLP and T2; D is the discriminator, and is binary cross entropy applied to a sigmoid output with binary class labels ; G is the generator; θ_G denotes the trainable parameters (i.e., weights) of the generator; $v_1, v_2, \ldots$ are input variables; Y is the precipitation at time grid cell i; and $\overline{Y}_i$ is the regional average precipitation. Based on a hyperparameter study, we selected values of 1 and 5 for w_a and w_c, respectively. We used Binary Cross Entropy as the loss function of the discriminator. Therefore, in addition to the L1-norm loss used to retain low-frequency content in the images, our target generator is trained to generate high-resolution precipitation patterns that are indistinguishable from the real high-resolution precipitation patterns generated by the discriminator. Once trained, only the generator is used for downscaling low-resolution precipitation data. We incorporate CGAN into both Direct-Simple and Encoded-Simple to obtain two new models that we refer to respectively as Direct-CGAN and Encoded-CGAN hereafter. We also use CGAN (with actual precipitation amounts) for training the SR model, producing a model described earlier and that we refer to as SR-CGAN.

2.6 Implementation and Model Training

We implement our model with the PyTorch machine learning framework. We use 3-hourly data from January to September of the year 2005 for model training and validation (e.g., hyperparameter tuning to control overfitting) and the remaining three months (October, November, December) for testing the model performance. The Adam optimizer and one NVIDIA V100 GPU are used for model training. The training is computationally intensive, taking for example about 4 hours for Encoded-CGAN using one NVIDIA V100 GPU for 8,000 iterations and a minibatch size of 32. Training time can be reduced by parallelism, for example to less than an hour on 8 GPUs. Once the model is trained, it takes less than a minute on one GPU to downscale three months of coarse-resolution precipitation data. Both training and inference for downscaling can be further accelerated using purposely built AI systems (Liu et al., 2021; Abeykoon et al., 2019) when deal with large datasets, e.g., downscaling variables of the Energy Exascale Earth System Model (E3SM Project, 2018, 2020) for higher spatial–temporal resolution.

2.7 Evaluation Metrics

We compare the 12 km precipitation field generated by the CNN models against two different sets of WRF precipitation outputs: one from WRF run at a grid spacing of 12 km (referred to as *Ground Truth*), and a second generated by interpolating output from a 50 km run to 12 km (referred to as *Interpolator*). We use the 12 km WRF modeled precipitation as Ground Truth because the CNNs are designed to achieve the performance of these data by approximating the relationship between coarse- and fine-resolution precipitation data. We examine the statistical distribution of precipitation using MSE and the probability density function (PDF) by aggregating all the grid cells over CONUS and smaller regions. MSE is computed for each timestep for the testing period:

$$\ell_{mse} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i^H - G_\theta\left(Y_i^L\right)\right)^2, \tag{2}$$

where N is the total number of grid cells over the study domain, Y_i^H and Y_i^L are the precipitation at grid cell (i) simulated by WRF at high (12 km) and low (50 km) resolution respectively. Y_i^H is used as Ground Truth in this case. G_θ is the deep neural network parameterized with θ that models the relationship between low and high resolution simulations. We calculate the MSE across CONUS and also in each of seven subregions defined by the National Climate Assessment (NCA; Melillo et al., 2014), as shown in Figure 6 (lower right).

To measure the similarity between the PDFs of Ground Truth, Interpolator, and the CNN models, we employ the Jensen–Shannon (J-S) distance (Osterreicher and Vajda, 2003; Endres and Schindelin, 2003), which measures the distance between two probability distributions. The J-S distance is computed by:

$$JSD(P,Q) = \sqrt{\frac{D(P||M) + D(Q||M)))}{2}}, \quad (3)$$

where P and Q are the two probability distributions to be evaluated (i.e., distribution of RCM simulated 12 km and CNN-downscaled 12 km precipitation, respectively), M is the mean of P and Q, and D is the Kullback–Leibler divergence (Kullback and Leibler, 1951) calculated with

$$D(P||M) = \sum_{x \in X} P(x) log \left(\frac{P(x)}{M(x)} \right), \quad (4)$$

where x is the bin we apply for the PDFs; here x = 0, 1, 2, ..., 30. We calculated the J-S distance (ranges from 0 to 1) between PDFs of the five CNN predictions and Ground Truth, with small (large) distance indicating that the CNN predictions have similar (different) distributions to Ground Truth. The J-S distance takes into account not only the median or the mean but also the entire distribution including the scale and the tails, which is important because variability changes are equally important especially for threshold-defined extremes, whose frequency is more sensitive to changes (Vitart et al., 2012).

We also investigate the geospatial pattern of J-S distance, as well as mean, standard deviation, and extreme values of precipitation over time. To measure whether the CNN models capture the spatial variability in these values of Ground Truth, we calculate the pattern correlations between each pair of the data, namely, Ground Truth versus Interpolator and Ground Truth versus each of the CNN models. Higher correlation indicates that the spatial variability in Ground Truth due to local effects (e.g., terrain) and synoptic-scale circulations are captured well by Interpolator and the CNN models.

While these metrics examine precipitation either at the level of individual model grid cells or at a regional scale by aggregating all the information together into one metric, they cannot determine model performance in rainstorm characteristics such as frequency, duration, intensity, and size of individual events. This information is all confounded in local or spatially aggregated time series. To overcome this limitation, we use a feature-tracking algorithm developed for rainstorm objects in particular (fully described in Chang et al., 2016), and we identify events using information from the precipitation field only. The algorithm applies almost-connected component labeling in a four-step process to reduce the influence of the chaining effect and allow grouping of physically reasonable events. The algorithm accounts for splits in individual events during their evolution and does not require that all precipitation be contiguous. In principle, such methods can decompose precipitation bias and distinguish between biases in the duration, intensity, size, and number of events. The duration of each event, in units of timesteps, is

$$D = T_e - T_b + 1, \quad (5)$$

where T_b and T_e are the beginning and ending timestep, respectively. The lifetime mean size S_{life} for an individual precipitation event is calculated as the sum of all the area (in km^2; derived by number of grid cells$\times 144$) associated with an event over its lifetime divided by D. The lifetime mean precipitation intensity is calculated by

$$I_{life} = \frac{V_{tot}}{S_{life}}, \tag{6}$$

where V_{tot} is the total precipitation volume (in m^3; derived by amount×144000) over the event lifetime.

3 Results

We now evaluate the efficacy of the five CNN methods by comparing their predictions with the original WRF output at a grid spacing of 12 km (Ground Truth), the output of a 12 km bilinear interpolation from the 50 km data (Interpolator), and the state-of-the-art SR-CGAN model output. We expect some of the CNN models developed by this study to generate more accurate results than Interpolator and SR-CGAN when using the same coarse-resolution RCM-modeled precipitation as input, because our CNN models are developed to approximate the relationship between the coarse- and the fine-resolution precipitation data.

3.1 MSE

Table 2 summarizes the MSE comparing the CNN-predicted precipitation with Ground Truth and Interpolator for the 50th and 99th percentiles picked from the testing period. The SR-CGAN model in general shows similar MSE to that of Interpolator, while the Simple models show smaller MSEs than that of Interpolator over several regions especially for 99th percentile of the precipitation. There are many timesteps and regions for which Direct-CGAN and Encoded-CGAN show larger MSEs than Interpolator and Simple models. The reason is that the Simple models are trained specifically to optimize this content-based loss, resulting in fields that are safer——that is, overly smoothed predictions of high-resolution precipitation. The CGAN models, in contrast, change the landscape of the loss function by adding the adversarial term to the L1-norm (Equation 1), which is more physically consistent with the training data, and by inserting significantly more small-scale features that better represent the nature of the true precipitation fields. However, these features also cause the high-resolution fields to deviate from Ground Truth in an MSE sense since they cannot be inferred from the low-resolution input.

3.2 Probability Density Function

To better validate that Direct-CGAN and Encoded-CGAN have learned the appropriate distribution for the precipitation data, we assess the PDFs of precipitation for the five CNN models across all timesteps over the entire CONUS and the seven subregions. For a certain region, we take into account all the grid cells and the timesteps (without any averages in space or time) for the density function. As shown in Figure 6, Ground Truth usually has longer tails than Interpolator and SR-CGAN have. SR-CGAN shows almost identical distribution to Interpolator, with smaller densities for the large precipitation than Ground Truth, and larger J-S distance from Ground Truth compared with the four new CNN models developed here. According to the J-S distance (Table 3), Direct-Simple and Encoded-Simple show closer distributions to Ground Truth than Interpolator and SR-CGAN do over some subregions such as Southwest, Northeast, Southern Great Plains, and Northwest, but they are still similar to or even worse than Interpolator over other regions, underestimating the heavier precipitation over the Northern

Table 2. MSEs (mm/3hr), calculated across all grid cells over the entire CONUS and seven subregions (Equation 2), at the 50th and 99th percentiles picked from all timesteps. The MSEs of the CGAN models are not necessarily smaller than those of the Simple models, especially for heavier precipitation, which has larger MSEs.

	CONUS	Southwest	Northeast	Midwest	S Great Pl.	Northwest	N Great Pl.	Southeast
Interpolator	0.242, 1.07	0.065, 2.26	0.038, 4.94	0.043, 4.35	0.004, 4.22	0.173, 3.3	0.062, 2.44	0.032, 4.41
SR-CGAN	0.251, 1.01	0.067, 2.13	0.033, 5.73	0.041, 4.49	0.004, 4.27	0.162, 3.01	0.065, 2.75	0.031, 5.4
Direct-Simple	0.201, 0.96	0.061, 1.45	0.044, 4.0	0.046, 3.82	0.007, 3.51	0.165, 2.75	0.052, 2.3	0.033, 4.27
Encoded-Simple	0.206, 1.03	0.058, 1.8	0.036, 4.57	0.04, 4.26	0.004, 4.73	0.171, 2.74	0.052, 2.29	0.033, 4.81
Direct-CGAN	0.21, 1.05	0.058, 1.93	0.036, 4.87	0.044, 4.81	0.003, 4.74	0.168, 3.27	0.055, 2.44	0.029, 4.95
Encoded-CGAN	0.223, 1.03	0.062, 1.87	0.037, 4.3	0.04, 4.62	0.003, 5.4	0.155, 3.02	0.055, 2.86	0.029, 5.79

Great Plains, Midwest, and Southeast. Direct-CGAN and Encoded-CGAN produce better precipitation distributions over these three subregions and show smaller J-S distance from Ground Truth than Interpolator, SR-CGAN, and the Simple models. This finding indicates that although Direct-Simple and Encoded-Simple obtain lower grid cell-wise error than Direct-CGAN and Encoded-CGAN do, they cannot capture the small-scale features (i.e., local extremes in time and space) that better represent the nature of the precipitation fields. In fact, we also investigate the differences in the lower part of the PDFs, and find that Direct-CGAN and Encoded-CGAN are closer to Ground Truth than Interpolator, SR-CGAN, as well as the Simple models over many subregions.

3.3 Geospatial Analysis of Other Measures

To investigate whether the five CNNs can capture the geospatial pattern of PDF distributions, mean, standard deviation, and top 5% precipitation seen in the Ground Truth, we conduct geospatial evaluations of these measures because they assess the performance of CNNs in a more accurate manner. For example, if the location of a heavy precipitation event is misrepresented in the CNN predictions, bias can be seen in geospatial maps but not in the PDF plot for a specific region (Figure 6).

First, we present J-S distance maps which quantify the similarity of the PDFs (based on time series) over each grid cell between Ground Truth and the six predictive models. As shown in Figure 7, Interpolator and SR-CGAN show similar J-S distance from Ground Truth with larger values over Northwest and Southeast. The Simple models show larger J-S distance from Ground Truth than Interpolator, SR-CGAN and the CGAN models do. Encoded-CGAN shows the smallest J-S distance from Ground Truth among the six predictive models, indicating it captures the precipitation distribution more accurately than other approaches do.

Second, we compare the mean state of precipitation for each dataset over all the grid cells. As shown in Figure 8, Interpolator shows a smoother precipitation pattern than Ground Truth does, with underestimation of heavy precipitation over the very northwestern part of CONUS, and overestimation of relatively light precipitation over other regions. SR-CGAN shows almost the same spatial pattern as Interpolator does. The fine-scale features added by SR-CGAN do not seem to help improve the

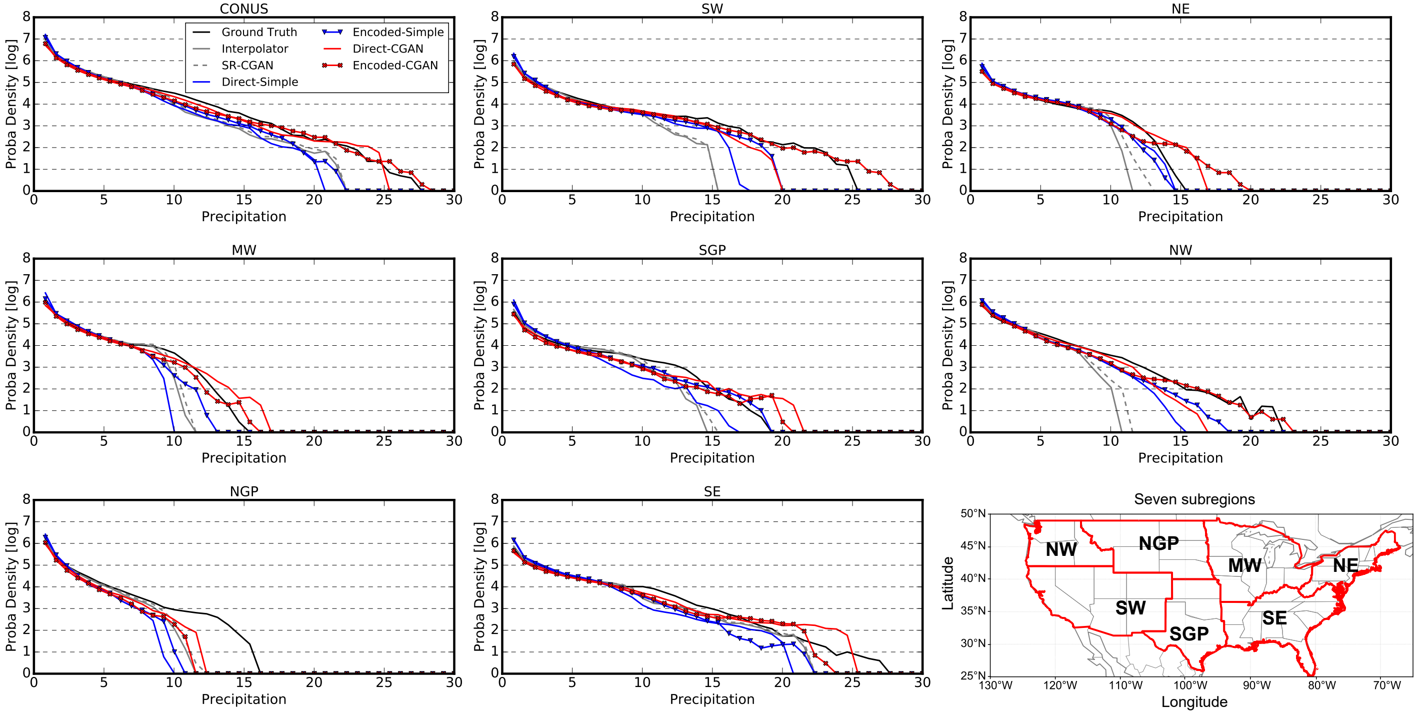


Figure 6. PDFs from Ground Truth, Interpolator, and CNN-predicted precipitation calculated based on grid cells and timesteps over CONUS and seven subregions. The subregions are based on those used in the national climate assessment: Northeast (NE), Southeast (SE), Midwest (MW), Southwest (SW), Northwest (NW), and Northern Great Plains (NGP), and Southern Great Plains (SGP).

underestimation of precipitation over the very Northwest. The pattern correlations between Ground Truth and Interpolator and between Ground Truth and SR-CGAN are also similar (0.866 and 0.871). The Simple models generate smaller-scale features, which are especially seen over the very Northwest, and the bias in precipitation along the western coast are smaller than in Interpolator and SR-CGAN. The pattern correlation between Ground Truth and Direct-Simple and that between Ground Truth and Encoded-Simple are improvements over Interpolator from 0.866 to 0.937 and 0.945, respectively. However, there is still an underestimation of high values (>1.7 mm) over Northwest. The precipitation patterns over the central and eastern United States generated by the Simple models are similar to those of Interpolator and SR-CGAN, with overestimation of light precipitation (0.1–0.5 mm) over southeastern and northeastern states such as Louisiana and Mississippi. The Direct-CGAN shows smaller precipitation bias than the Simple models do over Northwest; and both Direct-CGAN and Encoded-CGAN show smaller precipitation bias than Interpolator, SR-CGAN, and Simple models do over Great Plains and the southeastern states. The pattern correlation between Ground Truth and Direct-CGAN and that between Ground Truth and Encoded-CGAN are improved over Interpolator from 0.866 to 0.943 and 0.951, respectively. The improvements of the pattern correlation by

Table 3. J-S distance (Equation 3) measuring the similarity of the PDFs between Ground Truth and six predictive models (Interpolator and CNN-based models) over CONUS and seven subregions.

	CONUS	Southwest	Northeast	Midwest	S Great Plains	Northwest	N Great Plains	Southeast
Interpolator	0.164	0.325	0.241	0.229	0.210	0.372	0.275	0.151
SR-CGAN	0.162	0.325	0.174	0.219	0.187	0.348	0.255	0.150
Direct-Simple	0.200	0.275	0.069	0.293	0.149	0.247	0.336	0.188
Encoded-Simple	0.166	0.208	0.083	0.149	0.038	0.168	0.305	0.160
Direct-CGAN	0.081	0.209	0.138	0.130	0.152	0.202	0.239	0.103
Encoded-CGAN	0.039	0.107	0.187	0.069	0.115	0.060	0.271	0.125

the CGAN models indicate that they can better capture the spatial variability of the mean precipitation shown in Ground Truth than Interpolator and other CNN models do.

Next, we investigate the performance of the five CNN models for capturing higher-order statistics of precipitation, such as standard deviation and top 5% precipitation, since climate modeling is as much concerned with variability as with mean values. In Ground Truth the precipitation during October to December over the northwestern coast shows a standard deviation up to 2 mm/3 hr, the largest across the entire CONUS area. Standard deviation over the eastern United States is also large (1–1.7 mm), while that over Southwest is the smallest because this time period is usually very dry. All CNN models capture the geospatial pattern of standard deviation, with the largest value over the northwestern coast, followed by moderate value over the eastern United States and the smallest value over Southwest. However, as shown in Figure 9, Interpolator and SR-CGAN show similar patterns (with pattern correlations with Ground Truth of 0.845 and 0.836, respectively) and underestimate the precipitation variability over the very northwestern coast, especially along the western coast of Washington and Oregon. The four new CNN models developed in this study improve the standard deviation over Northwest. The pattern correlation between Ground Truth and four CNN models are also improved to greater than 0.9 (Table 4), indicating that the new CNN models capture not only the precipitation variability along the time over each grid cell but also the spatial variability of the standard deviation.

Last, we evaluate the model performance in the top 5% precipitation (averaged across 95th percentile to the maximum) during the testing period, the northwestern coast of Washington and Oregon and northern California have heavy precipitation, up to 10 mm/3 hr, and some southern states as well as the East Coast have precipitation up to 7 mm/3 hr. As shown by Figure 10, Interpolator and SR-CGAN underestimate the large precipitation over both the northwestern and the eastern coast of United States; but they overestimate the relatively small precipitation over several southern states, indicating a spatially smooth precipitation pattern . All four CNN models developed here improve the precipitation amount over northern California and over western Oregon and Washington. The spatial variability of the top 5% precipitation is also improved by the four new CNN models, with pattern correlation increases from 0.861 by Interpolator to 0.92–93 by the new CNN models.

We also study the geospatial pattern of the 70th to 99th percentiles of all the CNN-predicted precipitation by comparing with Ground Truth and Interpolator. The new CNNs consistently outperform Interpolator and SR-CGAN and perform well for

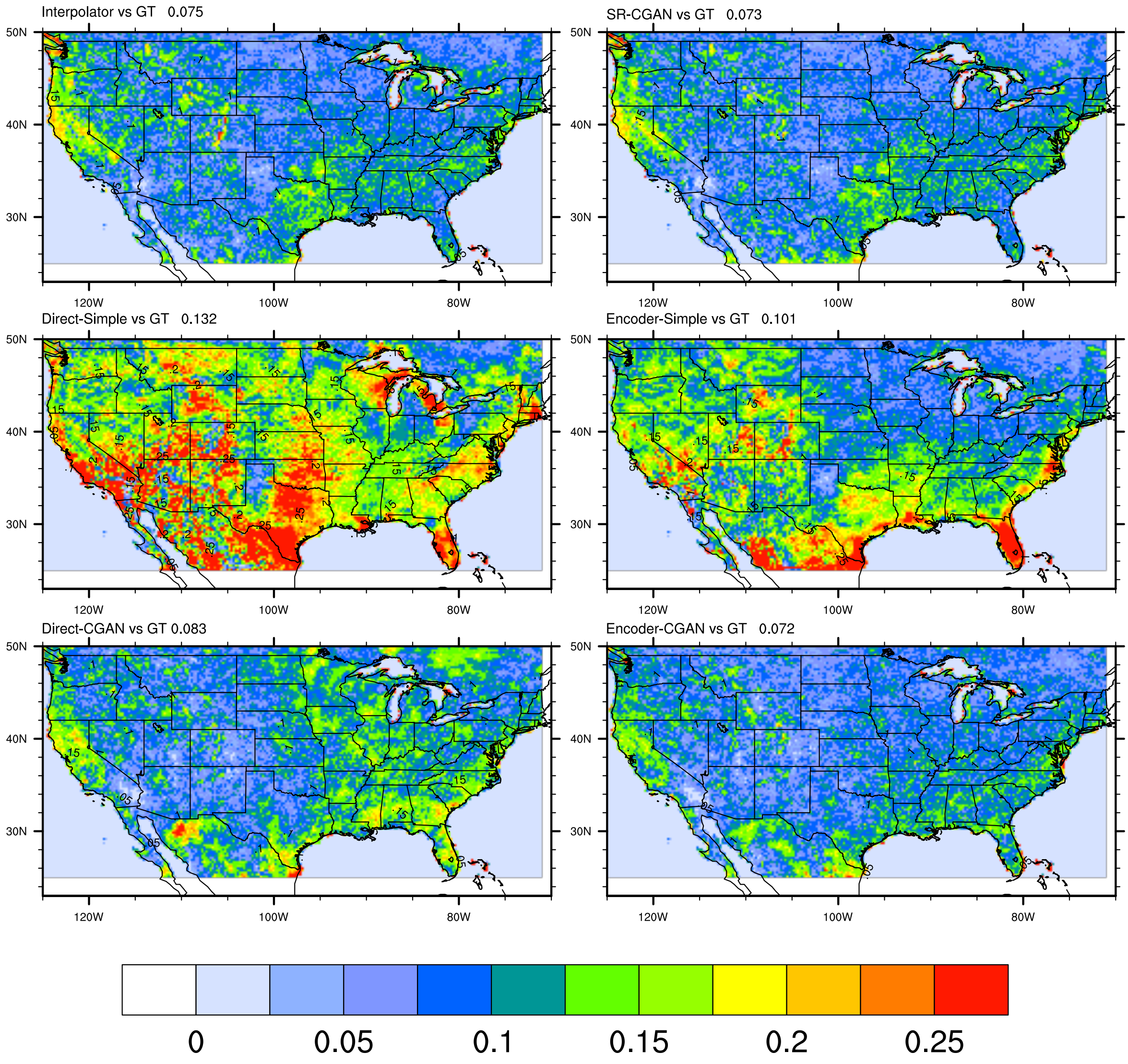


Figure 7. J-S distance (Equation 3) measuring the similarity of the PDFs between Ground Truth and six predictive models for the testing period (October–December).

increasingly extreme precipitation events, except for the most extreme (>97th). Compared with Interpolator, the Simple models clearly reduce the MSE and improve the pattern correlations, especially for percentiles higher than the 80th. In particular, Encoded-Simple outperforms Direct-Simple. Direct-CGAN and Encoded-CGAN further improve the pattern correlations over the Simple models and show the best match to the Ground Truth in terms of the spatial variabilites of the extreme precipitation.

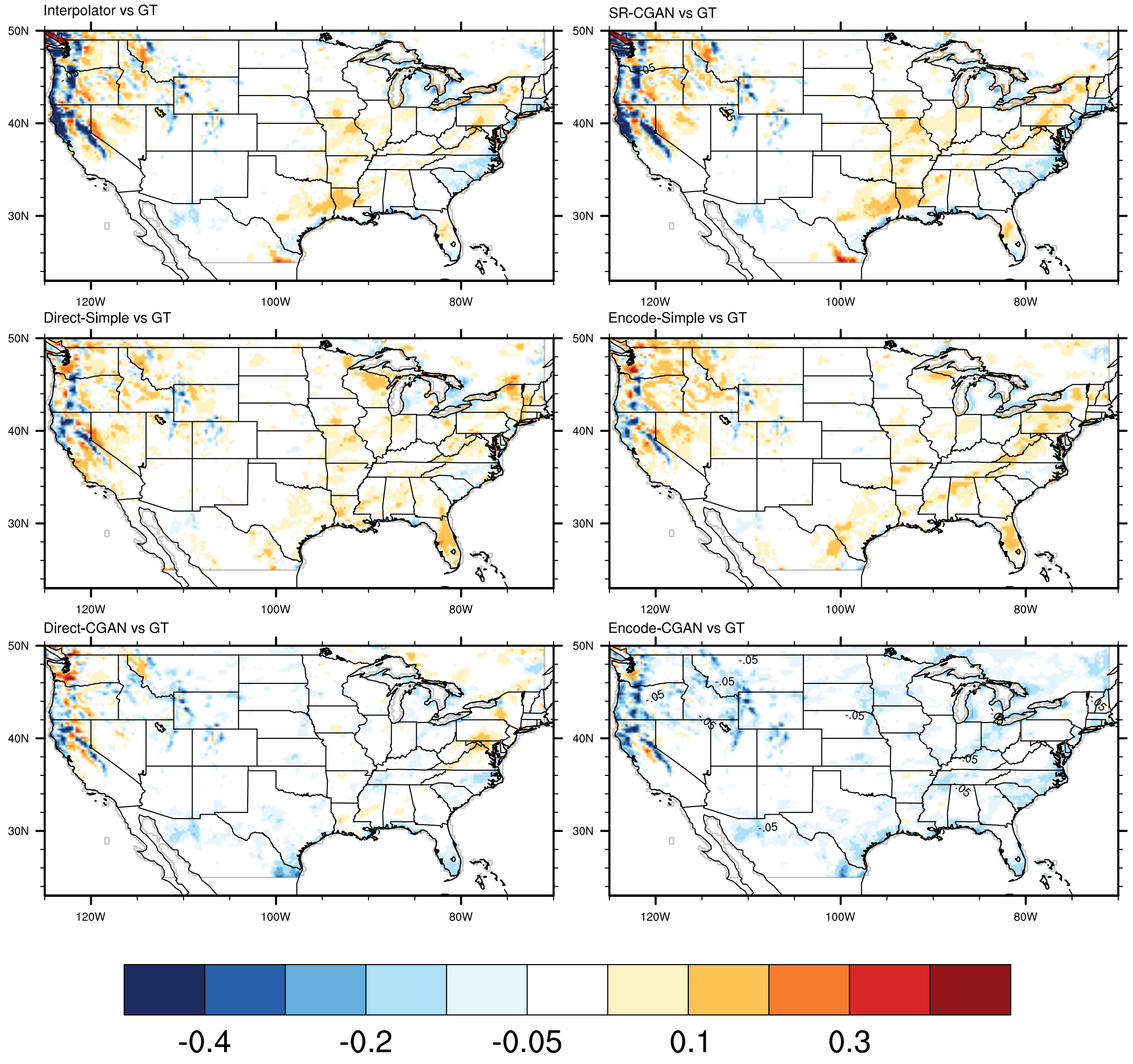


Figure 8. The differences (predictive models vs Ground Truth) of mean precipitation (mm/3 hr) produced by Ground Truth, Interpolator, and five CNN models for the testing period (October–December).

3.4 Event-Based Precipitation Characteristics

We investigate how well our different CNN models do at identifying and tracking precipitation events. To focus on events that are relevant to actual impact, we identify and track the precipitation events after removing grid cells with precipitation less than 10 mm/3 hr. First, we count the number of such events produced by each model during the testing period. We find 148 events in Ground Truth, 43 events in Interpolator, 57 events in SR-CGAN, 33 events in Direct-Simple, 69 events in Encoded-Simple,

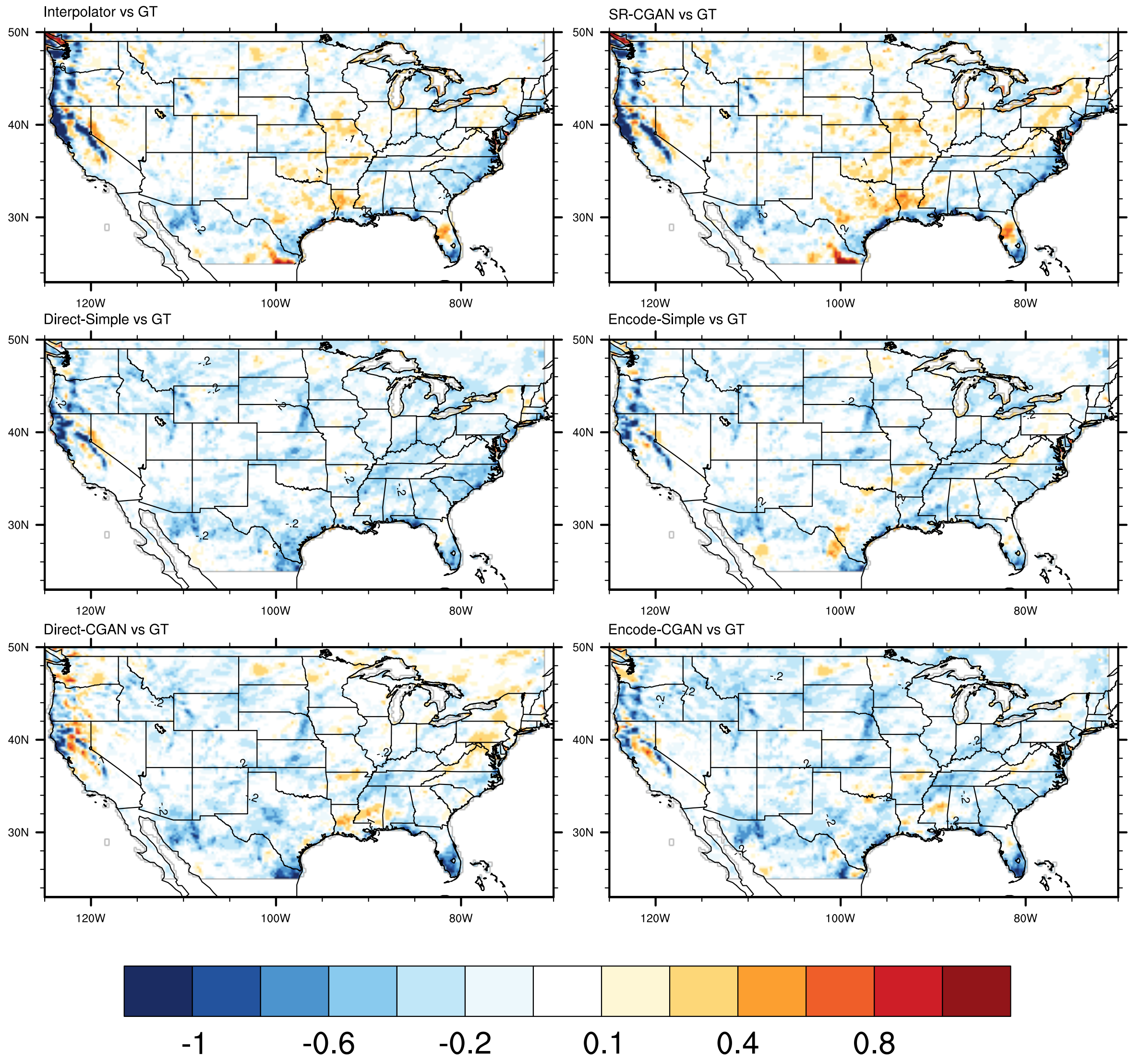


Figure 9. The differences (predictive models vs Ground Truth) in standard deviation of precipitation (mm/3 hr) produced by Ground Truth, Interpolator, and five CNN models for the testing period (October–December).

57 events in Direct-CGAN, and 84 events in Encoded-CGAN. Thus, while Encoded-CGAN does the best, all models greatly underestimate the number of events.

To examine how well different models capture the characteristics of the storms seen in Ground Truth, we track the life cycle of storm events in each CNN-prediction to determine the total volume, duration, lifetime mean size, and lifetime mean intensity of each event. We bin each of these characteristics and calculate their frequencies, giving the results in Figure 11. Looking first at precipitation intensity, we see that Ground Truth has intense precipitation events at greater than 11 mm/3 hr. While not

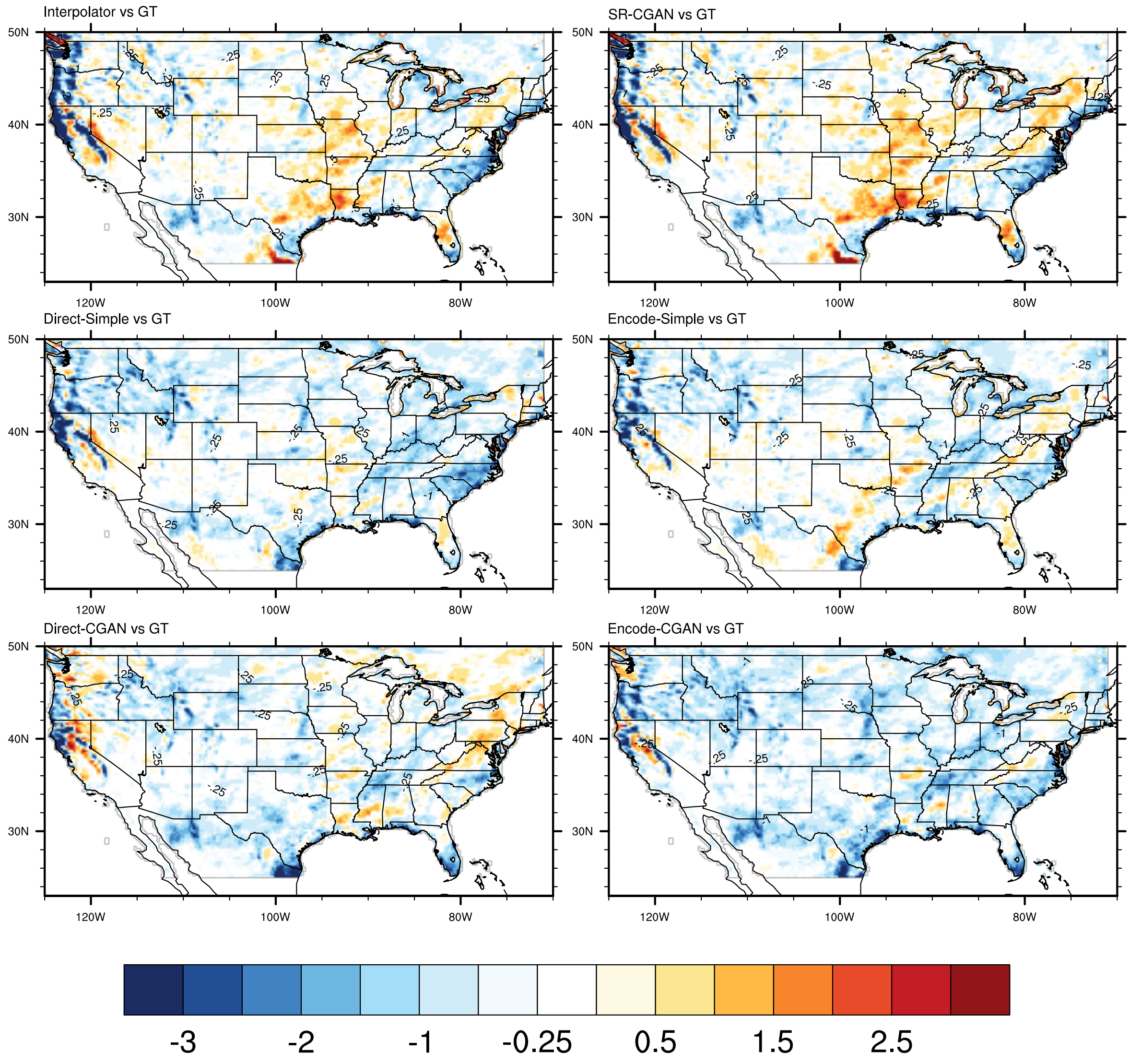


Figure 10. The differences (predictive models vs Ground Truth) in top 5% (averaged across the 95th percentile to maximum) of the precipitation amount (mm/3 hr) during the testing period (October–December).

captured by Interpolator, SR-CGAN, or Direct-Simple, this phenomenon is captured by Encoded-Simple, and Encoded-CGAN. This finding again indicates that the new CNN-based downscaling approach, especially when trained by CGAN, is useful for generating intense precipitation, while Interpolator and the state-of-the-art SR-CGAN tend to generate weaker precipitation events and cannot capture these strong events sufficiently.

Looking next at duration, we see that while all the models capture the most frequent short-term events (0–3 hours), SR-CGAN and Direct-Simple are not able to capture the longer-term events. For lifetime mean size, the new CNN models tend to

Table 4. Pattern correlations between Ground Truth and six predictive models (Interpolator and the CNN-based models).

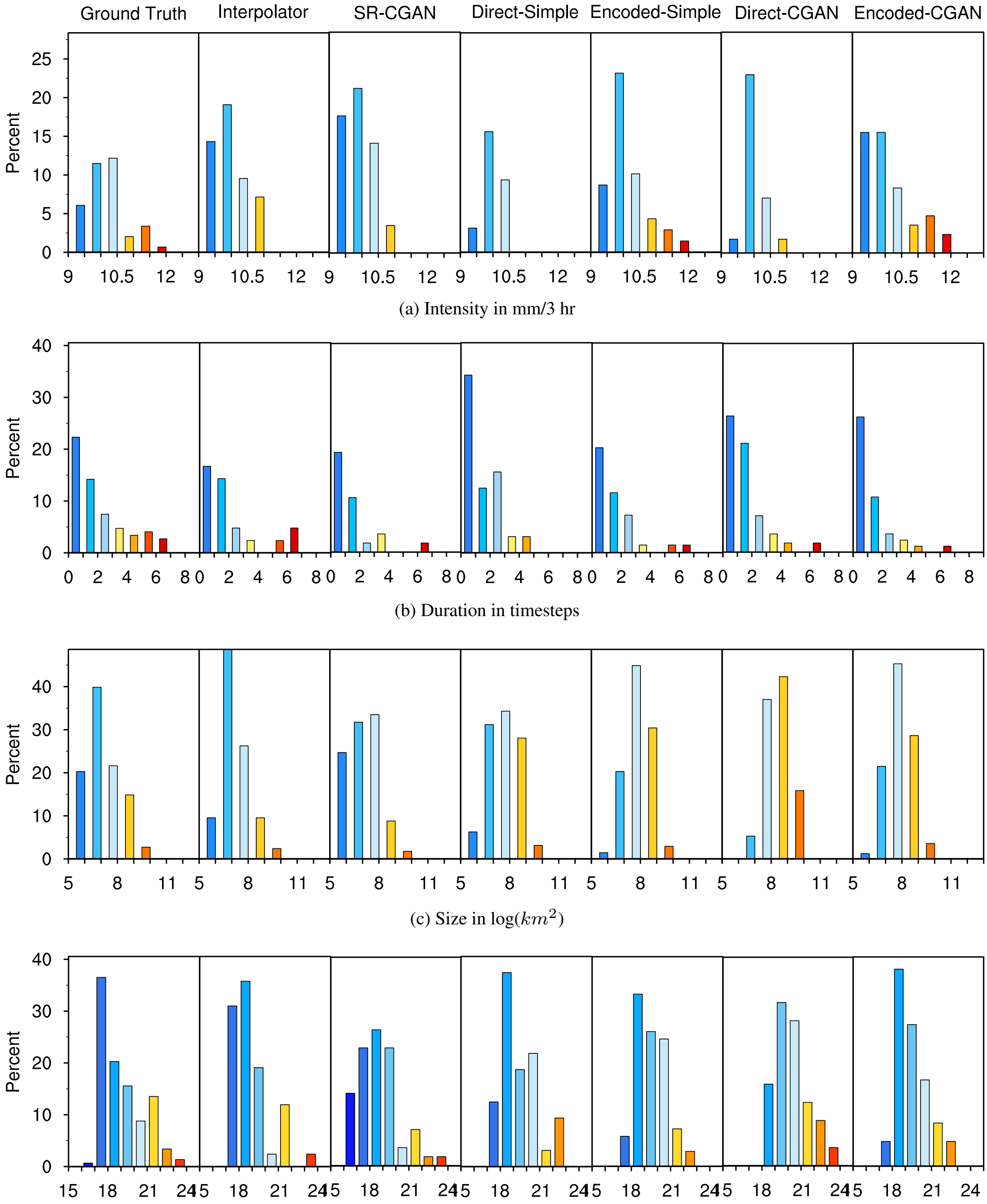
	Mean	Std. Dev	Top 5
Interpolator	0.866	0.845	0.861
SR-CGAN	0.871	0.836	0.854
Direct-Simple	0.937	0.915	0.928
Encoded-Simple	0.945	0.916	0.929
Direct-CGAN	0.943	0.903	0.922
Encoded-CGAN	0.951	0.912	0.930

produce more larger events but fewer smaller events than are seen in Ground Truth. Looking at total volume of precipitation, we see that because the four CNN models tend to have more intense, larger, and longer-duration events than Interpolator and SR-CGAN have, they show larger precipitation volumes more frequently than do Interpolator and SR-CGAN, and overall they better capture the large-volume precipitation events in Ground Truth. However, the four new CNN models—and, in particular, the CGAN models—do not show as frequent smaller-volume precipitation events as seen in Ground Truth. These small-volume precipitation events are well captured in Interpolator and SR-CGAN.

4 Summary and Discussion

This study develops a new CNN-based approach for downscaling precipitation from coarse spatial resolution RCMs. The downscaling approach is not constrained by the availability of observational data and can be applied to coarse-resolution simulation outputs to generate high-resolution precipitation maps with statistical properties (e.g., quantiles, including extremes and data variability) comparable to those seen in high-resolution RCM outputs. Because both the coarse-resolution simulations and neural network inferences are relatively inexpensive, our approach can greatly accelerate the process of generating such simulated high-resolution precipitation maps.

Our approach is different from the super-resolution approach to downscaling taken by previous studies. Our CNNs are developed by using two datasets that are generated by two sets of RCM simulations run at different spatial resolutions. The resulting precipitation differ not only in resolution but also sometimes in geospatial patterns. One of the reasons is that the high resolution modeling handles the topographic and scale-dependent physical processes (e.g., resolved scale convection, boundary layer phenomena, and parameterized clouds) better than does the low resolution modeling. In addition, the slightly different model domain coverage and the significantly different computing timesteps can cause differences in precipitation fields. To mitigate the challenge of learning the relationship between the low- and high-resolution simulations when precipitation occurs in different locations, we use an inception module in the neural network to learn information from not only a target grid cell but also its surroundings (so-called receptive fields) in previous layers and generate data for that particular grid cell for the next



(a) Intensity in mm/3 hr

(b) Duration in timesteps

(c) Size in log(km^2)

(d) Total volume in log(m^3)

Figure 11. Relative frequency (as %) of certain event-based precipitation characteristics: (a) lifetime mean intensity, mm/3 hr; (b) duration in (3-hr) timesteps; (c) lifetime mean size, km^2 (x-axis is in log); and (d) total volume, m^3 (x-axis is in log) during event lifetime.

layer. In addition, we employ the CGAN framework, which can help the CNN generate more physically realistic small-scale features and sharp gradients for precipitation in space.

We compare the new CNN-derived precpitation with precpitation generated from a Interpolator, which simply performs bilinear interpolation from coarse- to fine-resolution data, and that from the state-of-the-art SR-CGAN against Ground Truth. While Interpolator and SR-CGAN perform similarly, they are not as accurate as any of the four CNN methods when compared with Ground Truth, with overly smooth spatial precipitation patterns and underestimation of heavy precipitation. In contrast, the two new CGAN-trained CNNs produce the desired heavy-tailed shape, contributed by more intense and longer-lasting precipitation events that are in much better agreement with Ground Truth. In particular, when the CNN encoder is applied to the input variables, the output more accurately captures the spatial variabilities. To investigate whether the improvement of Encoded models is due to the larger number of parameters or the model architecture, we conducted another experiment by adding one more inception box (so do encoder layers) to Direct-CGAN so that its parameters is similar as Encoded-CGAN. Results show that the MSE of this larger-sized Direct-CGAN is close to the original Direct-CGAN, indicating that even with a larger number of parameters, the performance of Direct-CGAN is still the same. Therefore, the improvement of Encoded-CGAN compared with Direct-CGAN over many subregions are most likely due to the neural network architecture not the larger number of parameters. The findings of this study suggest that simply interpolating the coarse resolution or even using the state-of-the-art SR technique to generate fine-resolution data cannot capture the statistical distribution of precipitation.

The capability of generating high-resolution precipitation by using the technique developed in this study immediately suggests several interesting uses. For example, we can apply the CNN models to the outputs from the North American Regional Climate Change Assessment Program (NARCCAP; Mearns et al., 2012) or North American Coordinated Regional Climate Downscaling Experiments (NA-CORDEX; Mearns et al., 2017), both of which comprise multimodel 50 km WRF ensembles, to generate high-resolution precipitation data for uncertainty quantification in future projections. We also believe that the CNN models can be used to downscale output from a different RCM than that used to train the models, if that other RCM uses similar principal governing equations for simulated precipitation.

Although the CNNs that we have described here show promise, several limitations remain to be addressed. For example, we train the CNN with just nine months of data and test on the other three months of the same year. If we can train the CNN with multiple years of data that more fully capture the interannual variability, the algorithm might perform more robustly when applied to a new dataset. On the other hand, if we conduct both training and testing for a different year (e.g., use first 9 months for training, and the rest for testing), we expect the conclusion will be similar to what we draw from this study. It is possible though, that we may end up with a slightly different set of hyperparameters that achieves the best results. This will require careful tuning and testing of the model development. Another limitation is that our current CNN architecture does not consider dependencies between timesteps; instead, it processes images for each timestep independently. Therefore, the CNN output cannot capture temporal (here, 3-hourly) variations in precipitation data. However, the CNNs do capture the overall data variability and extremes over each grid cell, with improvements compared with Interpolator. Thus, peaks occurring at certain times in Ground Truth may not be captured by the CNN predictions, but the CNNs may have peaks with similar magnitudes at other times. While this performance is not satisfied in weather forecasting, it is acceptable for climate-scale simulations.

In fact, in climate science the preference is to compare the statistical distribution of weather events (e.g., climate) rather than actual day-to-day weather. Nevertheless, time dependencies in data can be important; weather propagates in time, and ideally precipitation should not be treated independently.

Other studies have used a recurrent neural network structure (Leinonen et al., 2020) to permit generated outputs to evolve in time in a consistent manner, so that the GAN generator can model the time evolution of fields and the discriminator can evaluate the plausibility of image sequences rather than single images. The 3-hourly data that we use in this study are potentially too coarse to consider time dependencies: short-duration events may disappear between timesteps, and even for long-duration events, 3 hours may be too long to capture a smooth transition from one timestep to the next, as preferred by the learning process. The other challenge is that once the time dimension is considered, the matrix will be three dimensional, which requires significantly larger computer memory. We plan to explore higher time-frequency dynamical downscaling simulations on more advanced GPU machines.

Code and data availability. Source code is available at https://github.com/lzhengchun/dsgan. The data used in this study are available at http://doi.org/10.5281/zenodo.4298978

Author contributions. JW participated in the entire project by providing domain expertise and analyzing the results from the CNNs. ZL designed, developed, and conducted all deep learning experiments. WC conducted event-based analysis. IF, RK, and VRK proposed the idea of this project and provided high-level guidance and insight for the entire study.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357, and was supported by a Laboratory Directed Research and Development (LDRD) Program at Argonne National Laboratory through U.S. Department of Energy (DOE) contract DE-AC02-06CH11357.

References

Abeykoon, V., Liu, Z., Kettimuthu, R., Fox, G., and Foster, I.: Scientific image restoration anywhere, in: 2019 IEEE/ACM 1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP), pp. 8–13, IEEE, 2019.

Barrett, A. I., Wellmann, C., Seifert, A., Hoose, C., Vogel, B., and Kunz, M.: One step at a time: How model time step significantly affects convection-permitting simulations, Journal of Advances in Modeling Earth Systems, 11, 641–658, 2019.

Bretherton, C. S. and Khairoutdinov, M. F.: Convective self-aggregation feedbacks in near-global cloud-resolving simulations of an aquaplanet, Journal of Advances in Modeling Earth Systems, 7, 1765–1787, 2015.

Chang, W., Stein, M. L., Wang, J., Kotamarthi, V. R., and Moyer, E. J.: Changes in spatiotemporal precipitation patterns in changing climate conditions, Journal of Climate, 29, 8355–8376, 2016.

Chang, W., Wang, J., Marohnic, J., Kotamarthi, V. R., and Moyer, E. J.: Diagnosing added value of convection-permitting regional models using precipitation event identification and tracking, Climate Dynamics, 55, 175–192, 2020.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: The role of internal variability, Climate Dynamics, 38, 527–546, 2012.

Dong, C., Loy, C. C., He, K., and Tang, X.: Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, pp. 184–199, Springer, 2014.

E3SM Project: Energy Exascale Earth System Model (E3SM), [Computer Software] https://dx.doi.org/10.11578/E3SM/dc.20180418.36, https://doi.org/10.11578/E3SM/dc.20180418.36, https://dx.doi.org/10.11578/E3SM/dc.20180418.36, 2018.

E3SM Project, D.: Energy Exascale Earth System Model v1.0, [Computer Software] https://doi.org/10.11578/E3SM/dc.20180418.36, https://doi.org/10.11578/E3SM/dc.20180418.36, https://doi.org/10.11578/E3SM/dc.20180418.36, 2018.

E3SM Project, D.: Energy Exascale Earth System Model v1.2.1, [Computer Software] https://doi.org/10.11578/E3SM/dc.20210309.1, https://doi.org/10.11578/E3SM/dc.20210309.1, https://doi.org/10.11578/E3SM/dc.20210309.1, 2020.

Endres, D. M. and Schindelin, J. E.: A new metric for probability distributions, IEEE Transactions on Information Theory, 49, 1858–1860, https://doi.org/10.1109/TIT.2003.813506, 2003.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling, International Journal of Climatology, 27, 1547–1578, 2007.

Geiss, A. and Hardin, J. C.: Radar super resolution using a deep convolutional neural network, Journal of Atmospheric and Oceanic Technology, pp. 1–29, 2019.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial nets, in: Advances in Neural Information Processing Systems, pp. 2672–2680, 2014.

Gutowski Jr, W. J., Ullrich, P. A., Hall, A., Leung, L. R., O'Brien, T. A., Patricola, C. M., Arritt, R., Bukovsky, M., Calvin, K., Feng, Z., et al.: The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information, Bulletin of the American Meteorological Society, 101, E664–E683, 2020.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095–1108, 2009.

Heavens, N. G., Ward, D. S., and Natalie, M.: Studying and projecting climate change with earth system models, Nature Education Knowledge, 4, 4, 2013.

Kirchmeier-Young, M., Gillett, N., Zwiers, F., Cannon, A., and Anslow, F.: Attribution of the influence of human-induced climate change on an extreme fire season, Earth's Future, 7, 2–10, 2019.

Komurcu, M., Emanuel, K., Huber, M., and Acosta, R.: High-resolution climate projections for the northeastern United States Using dynamical downscaling at convection-permitting scales, Earth and Space Science, 5, 801–826, 2018.

Kullback, S. and Leibler, R. A.: On information and sufficiency, Ann. Math. Statist., 22, 79–86, https://doi.org/10.1214/aoms/1177729694, https://doi.org/10.1214/aoms/1177729694, 1951.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86, 2278–2324, 1998.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, 2015.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690, 2017.

Legates, D. R.: Climate models and their simulation of precipitation, Energy and Environment, 25, 1163–1175, 2014.

Leinonen, J., Nerini, D., and Berne, A.: Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network, arXiv preprint arXiv:2005.10374, 2020.

Liu, Z., Bicer, T., Kettimuthu, R., and Foster, I.: Deep learning accelerated light source experiments, in: 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS), pp. 20–28, IEEE, 2019.

Liu, Z., Bicer, T., Kettimuthu, R., Gursoy, D., De Carlo, F., and Foster, I.: TomoGAN: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion, JOSA A, 37, 422–434, 2020a.

Liu, Z., Sharma, H., Park, J.-S., Kenesei, P., Almer, J., Kettimuthu, R., and Foster, I.: BraggNN: Fast X-ray Bragg Peak Analysis Using Deep Learning, arXiv preprint arXiv:2008.08198, 2020b.

Liu, Z., Ali, A., Kenesei, P., Miceli, A., Sharma, H., Schwarz, N., Trujillo, D., Yoo, H., Coffee, R., Herbst, R., et al.: Bridge Data Center AI Systems with Edge Computing for Actionable Information Retrieval, arXiv preprint arXiv:2105.13967, 2021.

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Reviews of Geophysics, 48, 2010.

Mearns, L., McGinnis, S., Korytina, D., Arritt, R., Biner, S., Bukovsky, M., Chang, H., Christensen, O., Herzmann, D., Jiao, Y., et al.: The NA-CORDEX dataset, version 1.0. NCAR climate data gateway, boulder CO, 2017.

Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., Caya, D., Correia Jr, J., Flory, D., Gutowski, W., , Takle, E. S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A. M. B., Qian, Y., Roads, J., Sloan, L., and Snyder, M.: The North American regional climate change assessment program: Overview of phase I results, Bulletin of the American Meteorological Society, 93, 1337–1362, 2012.

Melillo, J. M., Richmond, T., and Yohe, G. W.: Climate change impacts in the United States: The Third National Climate Assessment, Tech. rep., U.S. Global Change Research Program, 2014.

Mezghani, A., Dobler, A., Benestad, R., Haugen, J. E., Parding, K. M., Piniewski, M., and Kundzewicz, Z. W.: Subsampling impact on the climate change signal over Poland based on simulations from statistical and dynamical downscaling, Journal of Applied Meteorology and Climatology, 58, 1061–1078, 2019.

Miguez-Macho, G., Stenchikov, G. L., and Robock, A.: Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations, Journal of Geophysical Research: Atmospheres, 109, 2004.

Mirza, M. and Osindero, S.: Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784, 2014.

Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H., and Tomita, H.: Deep moist atmospheric convection in a subkilometer global simulation, Geophysical Research Letters, 40, 4922–4926, 2013.

Osterreicher, F. and Vajda, I.: A new class of metric divergences on probability spaces and its applicability in statistics, Annals of the Institute of Statistical Mathematics, 55, 639–653, 2003.

Prein, A. F. and Giangrande, S.: Sensitivity of organized convective storms to model grid spacing in current and future climates, Tech. rep., Brookhaven National Laboratory, 2020.

Skamarock, W. C. and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, Journal of Computational Physics, 227, 3465–3485, 2008.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2, Tech. rep., National Center For Atmospheric Research Boulder, 2005.

Stengel, K., Glaws, A., Hettinger, D., and King, R. N.: Adversarial super-resolution of climatological wind and solar data, Proceedings of the National Academy of Sciences, 117, 16 805–16 815, 2020.

Stouffer, R., Eyring, V., Meehl, G., Bony, S., Senior, C., Stevens, B., and Taylor, K.: CMIP5 scientific gaps and recommendations for CMIP6, Bulletin of the American Meteorological Society, 98, 95–105, 2017.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 2015.

Tian, J. and Ma, K.-K.: A survey on super-resolution imaging, Signal, Image and Video Processing, 5, 329–342, 2011.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: DeepSD: Generating high resolution climate change projections through single image super-resolution, in: 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1663–1672, 2017.

Vitart, F., Robertson, A. W., and Anderson, D. L.: Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate, Bulletin of the World Meteorological Organization, 61, 23, 2012.

Wang, J., Swati, F., Stein, M. L., and Kotamarthi, V. R.: Model performance in spatiotemporal patterns of precipitation: New methods for identifying value added by a regional climate model, Journal of Geophysical Research: Atmospheres, 120, 1239–1259, 2015.

Wang, Y., Sivandran, G., and Bielicki, J. M.: The stationarity of two statistical downscaling methods for precipitation under different choices of cross-validation periods, International Journal of Climatology, 38, e330–e348, 2018.

Woo, S., Park, J., Lee, J.-Y., and So Kweon, I.: CBAM: Convolutional block attention module, in: European Conference on Computer Vision, pp. 3–19, 2018.

Yang, C.-Y., Ma, C., and Yang, M.-H.: Single-image super-resolution: A benchmark, in: European Conference on Computer Vision, pp. 372–386, Springer, 2014.

Yashiro, H., Kajikawa, Y., Miyamoto, Y., Yamaura, T., Yoshida, R., and Tomita, H.: Resolution dependence of the diurnal cycle of precipitation simulated by a global cloud-system resolving model, Scientific Online Letters on the Atmosphere, 12, 272–276, 2016.

Zobel, Z., Wang, J., Wuebbles, D. J., and Kotamarthi, V. R.: Analyses for high-resolution projections through the end of the 21st century for precipitation extremes over the United States, Earth's Future, 6, 1471–1490, 2018.

Government License