
CONDITIONAL EMULATION OF GLOBAL PRECIPITATION WITH GENERATIVE ADVERSARIAL NETWORKS

Anonymous authors

Paper under double-blind review

ABSTRACT

Climate models encode our knowledge of the Earth system, enabling research on the earth’s future climate under alternative assumptions of how human-driven climate forcings, especially greenhouse gas emissions, will evolve. One important use of climate models is to estimate the impacts of climate change on natural and societal systems under these different possible futures. Unfortunately, running many simulations on existing models is extremely computationally expensive. These computational demands are particularly problematic for characterizing extreme events, which are rare and thus demand numerous simulations in order to precisely estimate the relevant climate statistics. In this paper we propose an approach to generating realistic global precipitation requiring orders of magnitude less computation, using a conditional generative adversarial network (GAN) as an emulator of an Earth System Model (ESM). Specifically, we present a GAN that emulates daily precipitation output from a fully coupled ESM, conditioned on monthly mean values. The GAN is trained to produce spatio-temporal samples: 28 days of precipitation in a 92×144 regular grid discretizing the globe. We evaluate the generator by comparing generated and real distributions of precipitation metrics including average precipitation, average fraction of dry days, average dry spell length, and average precipitation above the 90th percentile, finding the generated samples to closely match those of real data, even when conditioned on climate scenarios never seen during training.

1 INTRODUCTION

Climate models encapsulate our best understanding of the Earth system, allowing research to be conducted on Earth climates under alternative assumptions of how human-driven climate forcings are going to evolve. An important application of climate models is to provide metrics of mean and extreme climate changes, particularly under these alternative future scenarios, as these quantities drive the impacts of climate on society and natural systems (Mora et al., 2018; Forzieri et al., 2018; Raymond et al., 2020). Furthermore, efforts in integrated modeling seek to “close the loop,” by having impacts on society feedback on societal conditions that drive emissions (Calvin & Bond-Lamberty, 2018) through changes in industrial activity, energy and land use. Because of the need to explore a wide range of alternative scenarios and other sources of uncertainties in a computationally efficient manner, climate models can only take us so far, as they require large computational resources, with a single simulation of the 21st century taking on the order of weeks on a supercomputer. The computational requirements expand considerably when attempting to characterize the changing statistics of rare events, which requires large amounts of data, in order to accurately separate the signal of changing extremes from the noise intrinsic to the climate system.

Climate model emulators address the computational drawbacks of expensive climate models. Trained on climate model output, emulators are less complex, data-driven tools that are often obtained through parametric fits such as regressions. These parametric fits, while less accurate than climate models, can produce values in fractions of a second on everyday computational resources. Their computational cost, when significant, is made upfront in the training phase. Traditionally, emulators like Pattern Scaling have been used to approximate average quantities, from decadal through monthly averages of precipitation (Santer et al., 1990; Castruccio et al., 2014; Holden et al., 2014; Tebaldi & Arblaster, 2014; Herger et al., 2015; Kravitz et al., 2017; Link et al., 2019; Beusch et al., 2020; Nath et al., 2021). Recently, the accuracy of some of these top-down techniques for rep-

representing the mean behavior of extremes has been documented (Tebaldi et al., 2020). approaches to emulation involve directly approximating metrics themselves, such as the mean behavior of the highest precipitation day of the year (hence their label as "top-down" approaches). Alternatively, a "bottom-up" approach tackles the emulation of the building blocks, such as daily precipitation itself, from which metrics like the highest precipitation day of the year, and many others, can be derived. Examples include creating information from stochastic weather generators (Semenov & Barrow, 1997; Kilsby et al., 2007; Fatichi et al., 2011), which rely on parameterizing the distribution of the weather variable, and then randomly sampling its realizations. Weather generators have been widely applied but they are usually developed for individual locations and specific applications.

Joint efforts between machine learning and climate science have recently started to tackle some of the most complex data-driven problems (Reichstein et al., 2019; Rolnick et al., 2019). Most applications have focused on bringing deep learning in aid of improved model forecasts, model parameterizations, or in substitution of climate models (Cohen et al., 2019; Grover et al., 2015; Ham et al., 2019; Shi et al., 2017; Jones, 2017; He et al., 2020; Schneider et al., 2017; Weber et al., 2020; Schmidt et al., 2020); of improved detection of signals, from extreme events to large scale patterns of anthropogenic changes amidst the internal noise of the climate system (Liu et al., 2016; Wang et al., 2019; Klemmer et al., 2019; Barnes et al., 2020; Toms et al., 2020; Wills et al., 2020); and of spatial in-filling in the case of fine-scale features that models would be too expensive, or plainly unable, to generate, or observations cannot cover (Kühnlein et al., 2014; Amato et al., 2020; Vandal et al., 2017; Stengel et al., 2020).

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), most often used for generating high quality images, have recently been employed for precipitation nowcasting (Ravuri et al., 2021) and generating spatial samples (Besombes et al., 2021) emulating outputs from the PLASIM (Lunkeit et al., 2011) general circulation model. In this work we propose a GAN-based approach to emulating global climate model output for different climate scenarios. Our GANs emulate daily precipitation output from a fully coupled Earth System Model (ESM). The presented GANs are trained to produce samples in the form of $T \times H \times W$ tensors, where T denotes the number of timesteps (days) and H and W are the spatial height and width, respectively, of a regular grid discretizing the globe. The end-goal is for these samples to be statistically indistinguishable from samples of the same dimension drawn from a state-of-the-art ESM. Crucially, our spatio-temporal samples are be conditioned on mean monthly precipitation maps, allowing us to target generation to specific climate scenarios. Our trained GAN can rapidly generate numerous realizations at a vastly reduced computational expense, compared to large ensembles of climate models (Kay et al., 2015; Lehner et al., 2020), which greatly aids in estimating the statistics of extreme events.

2 MODEL

Architecture Our model architectures are based upon the BigGAN (Brock et al., 2019), with some key changes, including the omission of the self-attention layers due to computational limitations. The generator consists of five "generator blocks," for a total of 20 3d-convolutional layers; in all but the first block, there is $2\times$ upsampling halfway through the block in the height and width dimensions, while the number of channels halves each block in the generator. The input to the first block is $C \times T \times H \times W = 256 \times 42 \times 6 \times 9$ and the output of the last block is $8 \times 42 \times 96 \times 144$. The last block is fed into a $1 \times 1 \times 1$ 3d-convolution to produce the final $1 \times 42 \times 96 \times 144$ map. Note that the time dimension is fixed throughout, and is not progressively upsampled; initial experiments suggested that this performs better. The discriminator architecture effectively mirrors the generator architecture and follows BigGAN, with a few exceptions. Batch normalization is not used, as suggested by Gulrajani et al. (2017), and we replace the final global sum pooling with a linear channel summation. The architectures are visualized in Appendix A.

Progressive Training We make use of progressive training; specifically, we initially train the first generator and discriminator block to convergence, so that they are capable of generating and discriminating, respectively, on samples with low spatial resolution. After the first spatial resolution block has converged, we proceed, block-by-block, to train the remaining blocks analogously. This process continues until the final blocks have converged. Each time a new block is added, the training undergoes a 2 epoch (16,536 update) "fading period." During this fading period, the real and generated samples are linearly interpolated between the (old) lower resolution and new higher res-

olution samples. After the fading period, all samples remain at the higher resolution until the block is trained, up to a maximum of 15 additional epochs. The fading period provides a smoother transition between spatial resolutions, and side-steps training instability early in a block’s training. While the Wasserstein GAN loss is popular for stability purposes, we find that using binary cross entropy combined with the removal of the BigGAN’s linear layer trains stably and avoids mode collapse.

Conditioning In order to generate realistic data under previously unseen scenarios, we need a mechanism to produce conditional generation. In this work, we opt to condition our daily precipitation on a monthly precipitation map (averaged over 28-day “months”). In practice, the monthly average maps on which we condition could be produced rapidly via low resolution ESM emulators. We add auxiliary input channels to the generator and a loss term to penalize samples which do not respect the provided conditioning.

The input to each generator 3d-convolution operation is augmented with 24 additional channels:

- 1 conditioning map channel
- 4 “look-ahead 1” conditioning map channels
- 16 “look-ahead 2” conditioning map channels
- 1 land-ocean mask
- 2 spherical coordinate positional encodings

The first condition map channel is obtained by downsampling the full resolution conditioning map to the block’s resolution. The “look-ahead” condition maps are designed to provide higher resolution information about conditioning in earlier blocks. Assuming a block has spatial resolution $H \times W$, the look-ahead 1 channels are obtained as follows: first, downsample the full conditioning map to $2H \times 2W$; second, tile the map into the $H \times W$ tiles, where each tile contains four pixels, third, create one $H \times W$ map using the upper-left pixel in each tile, another using the upper-right pixel in each tile, and so forth, yielding 4 maps total. Look-ahead 2 is analogous to look-ahead 1, except the downsampling is to $4H \times 4W$, and each tile contains 16 pixels, which form the 16 maps. The look-ahead features proved to be key in obtaining accurate conditioning for higher blocks. The input to the convolution operations in the discriminator, in contrast, is only augmented with the land-ocean mask and spherical coordinates.

To encourage the generator to properly utilize the conditioning maps, we introduce a conditioning loss that penalizes the Euclidean distance between the conditioning map and the average precipitation over the central 28 days of the generated 42 day sequence. Imposing this loss over the central 28 days only allows the generator some flexibility near the boundaries, and helps reduce edge effects in the generated data (note that the discriminator acts upon the full 42 day sequence). The conditioning map is normalized using the same normalization in Sec. 3. For training, we produce our conditioning maps by averaging 28-day sequences of daily precipitation training data. Once fully trained, the 42 day samples from the generator are immediately center cropped to 28 days; for evaluation, “samples” refers to these shorter, month-like units.

3 EXPERIMENTAL SETUP

Data Our model uses daily output from CESM1-CAM5 ESM (Kay et al., 2015). CESM1-CAM5 fully couples atmosphere, ocean, land, and sea ice components at a $\sim 1^\circ$ spatial (horizontal) resolution. This work focuses on daily precipitation; however, the approach is general and could be extended in future work to other variables, such as temperature and humidity. Our data consists of 10 realizations of historical daily precipitation, each spanning the years 1920 to 2005, as well as 10 realizations under the Coupled Model Intercomparison Project Phase 5 (CMIP5) RCP 8.5 scenario, each spanning the years 2006-2100 (Taylor et al., 2012). RCP stands for “Representative Concentration Pathway,” and RCP8.5 is the scenario with the greatest amount of climate change among the CMIP5 RCPs. We choose it deliberately to confront our model with the widest range of behavior expected in future daily precipitation, expecting that the emulator will be accurate in representing less extreme scenarios after being trained on the most extreme. To account for the skewed distribution of daily precipitation values (mm/day), we apply $\log(1 + x)$ normalization. We randomly split the 10 historical realizations into training (8), validation (1) and test (1) sets; the RCP 8.5 data is also split 8-1-1. To test the model’s generalization to previously unseen climate scenarios, we also utilize two realizations from the RCP 6.0 scenario, each spanning 2006-2100.

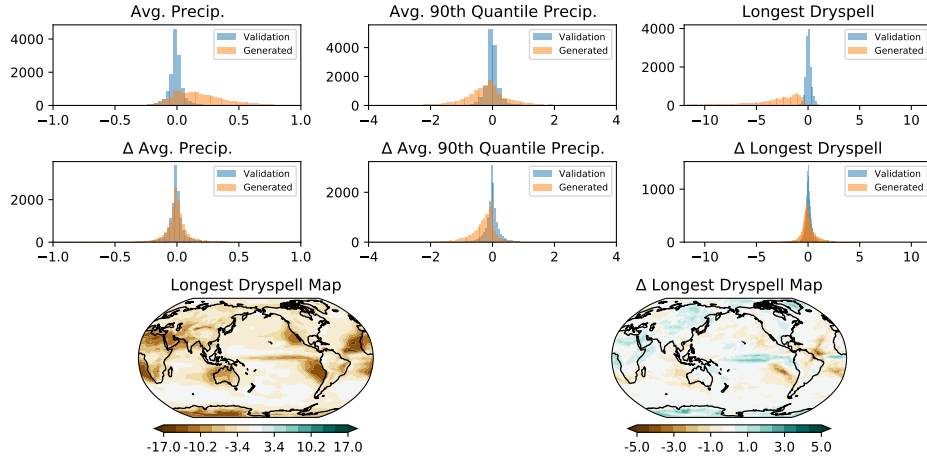


Figure 1: Top row: histograms over spatial locations, comparing differences between test and either validation or generated data, for three statistics. Second row: histograms over the differences with respect to the *change* in the metrics, relative to the reference period of 1960-1990. Third row: spatial distributions of errors for longest dry spell.

Evaluation Process Recall that the architecture’s goal is not to predict a particular outcome, but for the generator to define a distribution as close as possible to the distribution of the ESM output being emulated. We therefore compute a number of metrics (e.g., mean wet day precipitation, fraction of dry days, etc.) and their statistics over the 28-day samples from the GAN, and compare them to the same metrics/statistics computed over the ESM test data. The discrepancies between the two would ideally only reflect the noise from internal variability. To evaluate that, the discrepancies are compared to the corresponding ones computed between the validation data and the test data, both ESM output and therefore by construction only reflecting internal variability.

We first compute a suite of climate statistics over the monthly samples; specifically for each spatial location: (1) average precipitation, (2) average precipitation above the 90th quantile, (3) average number of dry days, (4) longest dryspell, (5) number of days with precipitation above the 90th quantile, and (6) the simple daily intensity index, SDII, which computes the average precipitation during wet days (Karl T.R. & A., 1999; T.C. & Coauthors, 2001). Then, for each of the $13824 = 96 \cdot 144$ spatial locations, we compute the average over all samples for each metric. We do this separately for the validation samples (from which conditioning maps were constructed), for the generated samples, and for the held out test set. Finally, we plot histograms of the differences between generated and test as well as the differences between validation and test. The validation-vs-test histogram serves as an upper bound on the expected agreement between our generated data and the test set.

Hyperparameter Tuning We use the AdamW optimizer (Loshchilov & Hutter, 2019), with β_1 and β_2 initialized to 0.0 and 0.9, respectively, and a batch size of 16. For the final block, at 96×144 resolution, we use higher β s of 0.9 and 0.999 for AdamW. The learning rate is set to $0.75e-4$ for both the generator and discriminator. The coefficient on the conditioning loss is set to 50.0. With no linear layer as described in BigGAN for the generator, our noise vector becomes $256 \times 42 \times 6 \times 9$ in size, resulting in 580,608 for initializing the first spatial resolution. The hyperparameters were minimally and manually tuned. Because the evaluation procedure described above involves a comparison between the validation and test sets for historical and RCP 8.5, we report our results on RCP 6.0 to avoid any potential bias that may have been introduced via tuning.

4 RESULTS AND DISCUSSION

The results for the average precipitation, average precipitation over the 90th quantile, and average dryspell length metrics are shown in Fig. 1; due to space constraints, the remaining metrics

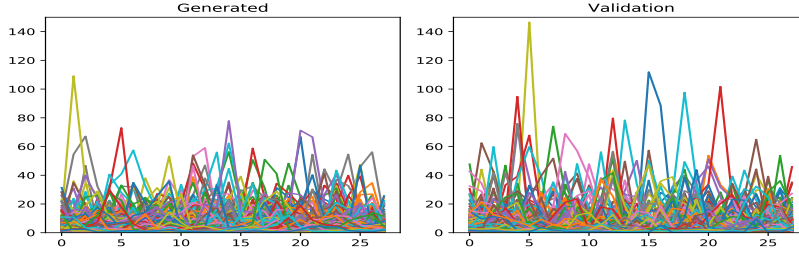


Figure 2: Generated (left) and validation (right) monthly samples for RCP6.0 2080-2100 for Hawaii.

are included as supplementary material in Appendix B. These plots all use the held out RCP6.0 emissions scenario over the time period 2006-2100. As described in Sec. 3, the histograms show validation-minus-test (blue) and generated-minus-test (orange), for each spatial location, averaged over samples. Because validation is (a) actual data from ESM and (b) the source of the conditioning for generation, we would expect the spread in orange to be equal or greater to blue. Among 5 of the 6 metrics, we see a small increase in the spread relative to validation, suggesting that GAN is approximating the ESM within a very small margin of error. The notable exception is the longest dry spell metric, which is strongly biased towards under-estimating this length. Interestingly, if we look at the *change* in these metrics relative to a 1960-1990 reference point, we see extremely good agreement (row 2). This suggests that the primary source of discrepancy is a spatial bias that persists across time, and not in the model’s ability to respond to the specific conditioning map. We investigate this further in the third row, which shows the spatial distribution of errors. Indeed, most of the errors occur over the ocean off of the west coast of land masses. We believe the generator is over-generalizing spatially, despite our use of spherical coordinates and land/ocean mask features, leading to errors where there are abrupt changes in dry spell behavior.

We also visualize the samples drawn from 2006-2100 for RCP6.0 for specific locations. Hawaii, USA is shown in Fig. 2, while Melbourne, Australia and Novosibirsk, Russia are included in Appendix C. We note that the samples appear to have similar spectral properties to the real data.

5 CONCLUSION AND FUTURE WORK

We have demonstrated the ability of a progressive conditional GAN architecture to successfully reproduce the spatio-temporal characteristics of sequences of daily precipitation produced by an ESM. With our approach, once trained, it becomes computationally trivial to produce new sample sequences, which is critical for the study of the risks of extreme weather events under hypothetical future emissions scenarios. We find the strongest model performance on average daily precipitation, and the weakest on estimating the length of the longest dry spell. We also find that most of the locations with the highest error are over the ocean, which minimizes their negative effects for some use cases, usually aimed at estimating impacts on land.

There are many ways we plan to extend the current work. The current work emulates only precipitation; adding support for multiple co-varying climate fields would not only enrich the potential use cases, but may result in a stronger generator that captures more of the underlying earth system dynamics. Established emulators of annual and monthly quantities on the basis of arbitrary scenarios of future forcing (Link et al., 2019; Nath et al., 2021) could be used here as the sources of conditional information. It would also be valuable to replicate these results across additional ESMs, and additional future climate scenarios. These capabilities, once developed, would support a significant improvement in integrated modeling of climate change impacts by enabling a rich representation of some of the most damaging hazards and an exploration of uncertainties in scenario, model and internal variability (Hawkins & Sutton, 2009; Lehner et al., 2020) “on the fly,” substituting for expensive, time consuming and necessarily constraining computational resources needed to run climate models.

ACKNOWLEDGMENTS

Redacted for anonymity.

REFERENCES

- Federico Amato, Fabian Guignard, Sylvain Robert, and Mikhail Kanevski. A novel framework for spatio-temporal prediction of climate data using deep learning. *CoRR*, 2020. URL <http://arxiv.org/abs/2007.11836>.
- Elizabeth A. Barnes, Benjamin Toms, James W. Hurrell, Imme Ebert-Uphoff, Chuck Anderson, and David Anderson. Indicator patterns of forced change learned by an artificial neural network. *Journal of Advances in Modeling Earth Systems*, 12(9), 2020. doi: 10.1029/2020MS002195.
- C. Besombes, O. Pannekoucke, C. Lapeyre, B. Sanderson, and O. Thual. Producing realistic climate data with generative adversarial networks. *Nonlinear Processes in Geophysics*, 28(3): 347–370, 2021. doi: 10.5194/npg-28-347-2021. URL <https://npg.copernicus.org/articles/28/347/2021/>.
- Lea Beusch, Lukas Gudmundsson, and Sonia I. Seneviratne. Emulating Earth system model temperatures with MESMER: from global mean temperature trajectories to grid-point-level realizations on land. *EARTH SYSTEM DYNAMICS*, 11(1):139–159, FEB 17 2020. ISSN 2190-4979. doi: {10.5194/esd-11-139-2020}.
- Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *CoRR*, 2019. URL <http://arxiv.org/abs/1809.11096>.
- Katherine Calvin and Ben Bond-Lamberty. Integrated human-earth system modeling—state of the science and future directions. *Environmental Research Letters*, 13(6):063006, jun 2018. doi: 10.1088/1748-9326/aac642. URL <https://doi.org/10.1088%2F1748-9326%2Faac642>.
- Stefano Castruccio, David J. McInerney, Michael L. Stein, Feifei Liu Crouch, Robert L. Jacob, and Elisabeth J. Moyer. Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs. *JOURNAL OF CLIMATE*, 27(5):1829–1844, MAR 2014. ISSN 0894-8755. doi: {10.1175/JCLI-D-13-00099.1}.
- Judah Cohen, Dim Coumou, Jessica Hwang, Lester Mackey, Paulo Orenstein, Sonja Totz, and Eli Tziperman. S2s reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *WIREs Climate Change*, 10(2):e00567, 2019. doi: 10.1002/wcc.567. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.567>.
- Simone Fatichi, Valeriy Y. Ivanov, and Enrica Caporali. Simulation of future climate scenarios with a weather generator. *Advances in Water Resources*, 34(4):448 – 467, 2011. ISSN 0309-1708. doi: <https://doi.org/10.1016/j.advwatres.2010.12.013>. URL <http://www.sciencedirect.com/science/article/pii/S0309170811000042>.
- G. Forzieri, A. Bianchi, F. Batista e Silva, M. A. Marin Herrera, A. Leblois, C. Lavalle, J.C.J.H. Aerts, and L. Feyen. Escalating impacts of climate extremes on critical infrastructures in europe. *Global Environmental Change*, 48:97 – 107, 2018. ISSN 0959-3780. doi: <https://doi.org/10.1016/j.gloenvcha.2017.11.007>.
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- Aditya Grover, Ashish Kapoor, and Eric Horvitz. A deep hybrid model for weather forecasting. In *Proc. KDD*, pp. 379–386, 2015. ISBN 9781450336642. doi: 10.1145/2783258.2783275.
- Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville. Improved training of wasserstein gans. *CoRR*, abs/1704.00028, 2017. URL <http://arxiv.org/abs/1704.00028>.
- Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. Deep learning for multi-year enso forecasts. *Nature*, 573(7775):568–572, Sep 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1559-7. URL <https://doi.org/10.1038/s41586-019-1559-7>.

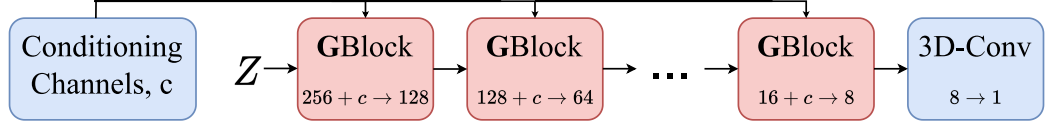
-
- E. Hawkins and R. Sutton. The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, 90(8):1095–1108, 2009. doi: 10.1175/2009BAMS2607.1.
- Sijie He, Xinyan Li, Timothy DelSole, Pradeep Ravikumar, and Arindam Banerjee. Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances. *CoRR*, 2020. URL <http://arxiv.org/abs/2006.07972>.
- N. Herger, B. M. Sanderson, and R. Knutti. Improved pattern scaling approaches for the use in climate impact studies. *Geophysical Research Letters*, 42(9):3486–3494, 2015. doi: 10.1002/2015GL063569.
- P. B. Holden, N. R. Edwards, P. H. Garthwaite, K. Fraedrich, F. Lunkeit, E. Kirk, M. Labriet, A. Kanudia, and F. Babonneau. PLASIM-ENTSem v1.0: a spatio-temporal emulator of future climate change for impacts assessment. *GEOSCIENTIFIC MODEL DEVELOPMENT*, 7(1):433–451, 2014. ISSN 1991-959X. doi: {10.5194/gmd-7-433-2014}.
- Nicola Jones. How machine learning could help to improve climate forecasts. *Nature*, 548(7668): 379–380, Aug 2017. ISSN 0028-0836. doi: <https://doi.org/10.1038/548379a>.
- Nicholls N. Karl T.R. and Ghazi A. Clivar/gcos/wmo workshop on indices and indicators for climate extremes: Workshop summary. *Climatic Change*, 42:3–7, 1999. doi: 10.1023/A:1005491526870.
- J. E. Kay, C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. M. Arblaster, S. C. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J.-F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein. The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349, 09 2015. ISSN 0003-0007. doi: 10.1175/BAMS-D-13-00255.1. URL <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- C.G. Kilsby, P.D. Jones, A. Burton, A.C. Ford, H.J. Fowler, C. Harpham, P. James, A. Smith, and R.L. Wilby. A daily weather generator for use in climate change studies. *Environmental Modelling & Software*, 22(12):1705 – 1719, 2007. ISSN 1364-8152. doi: <https://doi.org/10.1016/j.envsoft.2007.02.005>. URL <http://www.sciencedirect.com/science/article/pii/S136481520700031X>.
- Konstantin Klemmer, Adriano Koshiyama, and Sebastian Flennerhag. Augmenting correlation structures in spatial data using deep generative models. *CoRR*, 2019. URL <http://arxiv.org/abs/1905.09796>.
- B. Kravitz, C. Lynch, C. Hartin, and B. Bond-Lamberty. Exploring precipitation pattern scaling methodologies and robustness among cmip5 models. *Geoscientific Model Development*, 10(5):1889–1902, 2017. doi: 10.5194/gmd-10-1889-2017. URL <https://www.geosci-model-dev.net/10/1889/2017/>.
- Meike Kühnlein, Tim Appelhans, Boris Thies, and Thomas Nauss. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — a random forests-based approach applied to msg seviri. *Remote Sensing of Environment*, 141:129 – 143, 2014. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2013.10.026>. URL <http://www.sciencedirect.com/science/article/pii/S0034425713003945>.
- Flavio Lehner, Clara Deser, Nicola Maher, Jochem Marotzke, Erich M. Fischer, Lukas Brunner, Reto Knutti, and Ed Hawkins. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *EARTH SYSTEM DYNAMICS*, 11(2):491–508, MAY 29 2020. ISSN 2190-4979. doi: {10.5194/esd-11-491-2020}.
- R. Link, A. Snyder, C. Lynch, C. Hartin, B. Kravitz, and B. Bond-Lamberty. Fldgen v1. 0: an emulator with internal variability and space–time correlation for earth system models. *Geoscientific Model Development (Online)*, 12(4):1477–1489, 2019.

-
- Yunjie Liu, Evan Racah, Prabhat, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, and William Collins. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *CoRR*, 2016. URL <http://arxiv.org/abs/1605.01156>.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proc. ICLR*, 2019.
- Frank Lunkeit, Hartmut Borth, Michael Böttinger, Klaus Fraedrich, Heiko Jansen, Edilbert Kirk, Axel Kleidon, Ute Luksch, Pablo Paiewonsky, Silke Schubert, Frank Sielmann, and Hui Wan. Planet simulator reference manual version 16. 2011.
- C. Mora, D. Spirandelli, E.C. Franklin, J. Lynham, M.B. Kantar, W. Miles, C.Z. Smith, K. Freel, J. Moy, L. V. Louis, E. W. Barba, K. Bettinger, A.G. Frazier, J.F. Colburn IX, N. Hanasaki, E. Hawkins, Y. Hirabayashi, W. Knorr, C.M. Little, K. Emanuel, J. Sheffield, J.A. Patz, and C. L. Hunter. Broad threat to humanity from cumulative climate hazards intensified by greenhouse gas emissions. *Nature Climate Change*, 8:1062–1071, 2018. doi: 10.1038/s41558-018-0315-6.
- S. Nath, Q. Lejeune, L. Beusch, C.-F. Schleussner, and S. I. Seneviratne. Mesmer-m: an earth system model emulator for spatially resolved monthly temperatures. *Earth System Dynamics Discussions*, 2021:1–38, 2021. doi: 10.5194/esd-2021-59. URL <https://esd.copernicus.org/preprints/esd-2021-59/>.
- Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, Sep 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03854-z. URL <https://doi.org/10.1038/s41586-021-03854-z>.
- Colin Raymond, Radley M. Horton, Jakob Zscheischler, Olivia Martius, Amir AghaKouchak, Jennifer Balch, Steven G. Bowen, Suzana J. Camargo, Jeremy Hess, Kai Kornhuber, Michael Oppenheimer, Alex C. Ruane, Thomas Wahl, and Kathleen White. Understanding and managing connected extreme events. *Nature Climate Change*, 10(7):611–621, Jul 2020. ISSN 1758-6798. doi: 10.1038/s41558-020-0790-4. URL <https://doi.org/10.1038/s41558-020-0790-4>.
- Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, Feb 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-0912-1. URL <https://doi.org/10.1038/s41586-019-0912-1>.
- David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling climate change with machine learning. *CoRR*, abs/1906.05433, 2019. URL <http://arxiv.org/abs/1906.05433>.
- B. D. Santer, T. M. L. Wigley, M. E. Schlesinger, and J.F.B. Mitchell. Developing climate scenarios from equilibrium gcm results. *Report of the Max Plank Institut fur Meteorologie*, 47:29, 1990.
- Victor Schmidt, Mustafa Alghali, Kris Sankaran, Tianle Yuan, and Yoshua Bengio. Modeling cloud reflectance fields using conditional generative adversarial networks. *CoRR*, 2020. URL <http://arxiv.org/abs/2022.07579>.
- Tapio Schneider, Shiwei Lan, Andrew Stuart, and João Teixeira. Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, 44(24):12,396–12,417, 2017. doi: 10.1002/2017GL076101. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076101>.

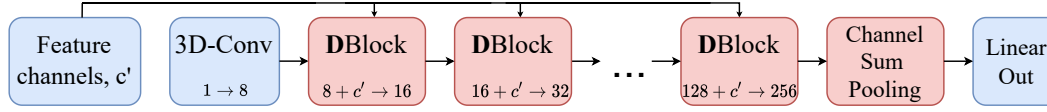
-
- Mikhail A. Semenov and Elaine M. Barrow. Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, 35(4):397–414, Apr 1997. ISSN 1573-1480. doi: 10.1023/A:1005342632279. URL <https://doi.org/10.1023/A:1005342632279>.
- Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Proc. NIPS*, pp. 5617–5627. 2017.
- Karen Stengel, Andrew Glaws, Dylan Hettinger, and Ryan N. King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815, 2020. ISSN 0027-8424. doi: 10.1073/pnas.1918964117.
- Karl E. Taylor, Ronald J. Stouffer, and Gerald A. Meehl. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498, 2012. doi: 10.1175/BAMS-D-11-00094.1. URL <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Peterson T.C. and Coauthors. Report on the activities of the working group on climate change detection and related rapporteurs. pp. 143, 2001. URL <http://etccdi.pacificclimate.org/docs/wgccd.2001.pdf>.
- C Tebaldi, A Armbruster, H P Engler, and R Link. Emulating climate extreme indices. *Environmental Research Letters*, 15(7):074006, jun 2020. doi: 10.1088/1748-9326/ab8332. URL <https://doi.org/10.1088/1748-9326/ab8332>.
- Claudia Tebaldi and Julie M. Arblaster. Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Climatic Change*, 122(3):459–471, Feb 2014. ISSN 1573-1480. doi: 10.1007/s10584-013-1032-9. URL <https://doi.org/10.1007/s10584-013-1032-9>.
- Benjamin A. Toms, Elizabeth A. Barnes, and Imme Ebert-Uphoff. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, 12(9), 2020. doi: 10.1029/2019MS002002.
- Thomas Vandal, Evan Kodra, Sangram Ganguly, Andrew R. Michaelis, Ramakrishna R. Nemani, and Auroop R. Ganguly. DeepSD: Generating high resolution climate change projections through single image super-resolution. *CoRR*, abs/1703.03126, 2017. URL <http://arxiv.org/abs/1703.03126>.
- Senzhang Wang, Jiannong Cao, and Philip S. Yu. Deep learning for spatio-temporal data mining: A survey. *CoRR*, 2019. URL <http://arxiv.org/abs/1906.04928>.
- T. Weber, A. Corotan, B. Hutchinson, B. Kravitz, and R. Link. Technical note: Deep learning for creating surrogate models of precipitation in earth system models. *Atmospheric Chemistry and Physics*, 20(4):2303–2317, 2020. doi: 10.5194/acp-20-2303-2020. URL <https://acp.copernicus.org/articles/20/2303/2020/>.
- Robert C. J. Wills, David S. Battisti, Kyle C. Armour, Tapio Schneider, and Clara Deser. Pattern Recognition Methods to Separate Forced Responses from Internal Variability in Climate Model Ensembles and Observations. *Journal of Climate*, 33(20):8693–8719, 09 2020. ISSN 0894-8755. doi: 10.1175/JCLI-D-19-0855.1. URL <https://doi.org/10.1175/JCLI-D-19-0855.1>.

A ARCHITECTURE FIGURES

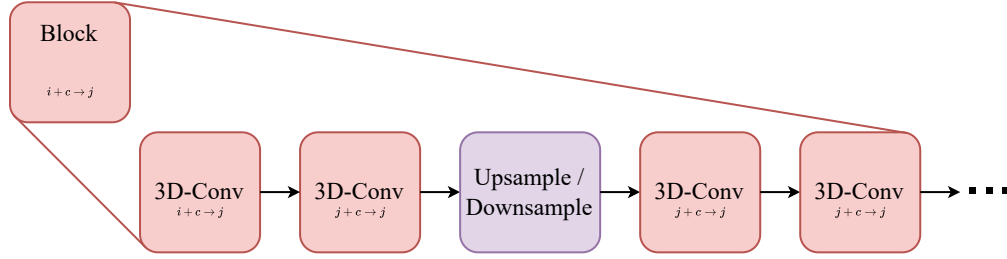
The generator and discriminator are visualized as a series of blocks in Figures 3a and 3b, respectively. The components of each generator and discriminator block are in turn shown in Fig. 3c. Note that the middle component of each generator block (except the first) spatially upsamples, while the middle component of each discriminator block downsamples. For this work, the number of additional input channels to the generator is $c = 24$, while the number of additional feature channels to the discriminator is $c' = 3$.



(a) Generator architecture



(b) Discriminator architecture



(c) Block details

Figure 3: Model Details

B ADDITIONAL METRICS

Fig. 4 visualizes results on the remaining three metrics. Again, we see reasonable performance on the metrics themselves, and quite good agreement as changes relative to the reference period.

C ADDITIONAL SAMPLES

Fig. 5 plots RCP 6.0 samples for Melbourne, Australia (top) and Novosibirsk, Russia (bottom). Melbourne’s precipitation is similar, with perhaps fewer high precipitation days in the generated samples as compared to the validation data. In Novosibirsk, the samples are again similar in nature; although here the outliers are larger in the generated data than the validation data.

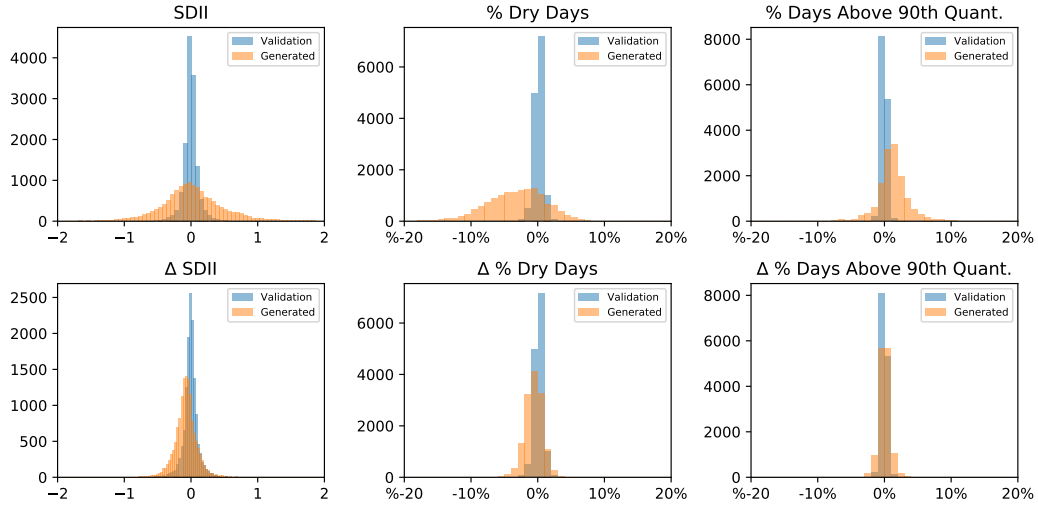


Figure 4: Top row: histograms over spatial locations, comparing differences between test and either validation or generated data, for three statistics. Second row: histograms over the differences with respect to the *change* in the metrics, relative to the reference period of 1960-1990.

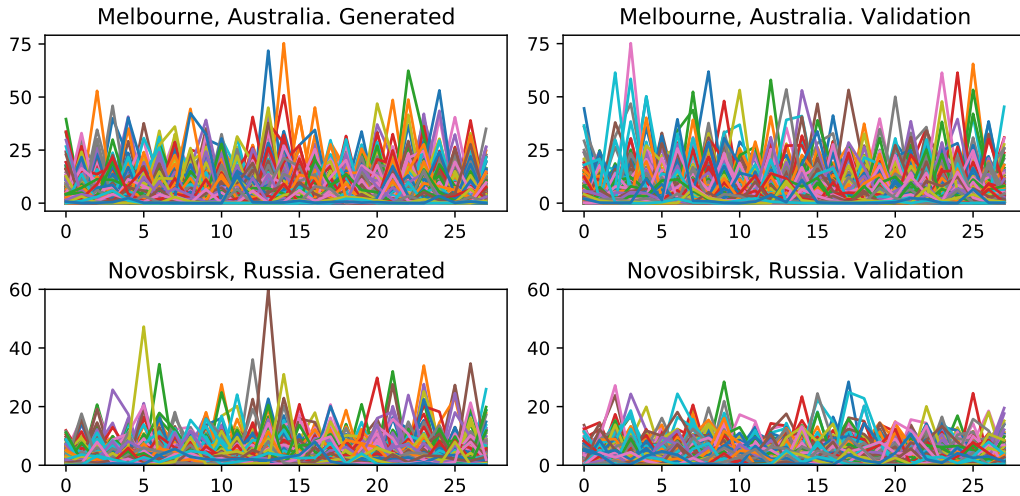


Figure 5: Generated (left) and validation (right) monthly samples for RCP6.0 2080-2100 for Melbourne, Australia (top row) and Novosibirsk, Russian (bottom row).