Downscaling using Deep Convolutional Autoencoders, a case study for South East Asia

September 2022

The manuscript under consideration leverages convolutional Autoencoders for both downscaling and bias correction in CMIP5 two-meter temperature data with higher resolution ERA5 data as a target. More specifically, the authors use three different approaches: (AE) simultaneous bias correction and downscaling, (ST) both bias and downscaling but in two stages with two autoencoders, and (RC) downscaling only. The authors examine CMIP5 and CMIP6 as well as a regional climate model over a similar geographic domain (Southeast Asia) to analyze the performance of the data-driven downscaling and bias correction. These convolutional autoencoders are able to super-resolve the temperature fields by a factor of five while preserving some physical structure. The authors use RCP pathways 2.6 and 4.5 as testbeds for the generalizability of the approach. Ultimately, they outline several directions for future work including increased precision and multi-variate approaches.

There are a number of aspects to this manuscript that make it appealing. It is a fairly novel use of Autoencoders, which have traditionally been used for representational learning rather than as generative modeling in the weather/climate space. And the authors are working to address what is an important limitation in the community (biases and low resolution of temperature, particularly with orography) that needs to be corrected. However, this manuscript has significant shortcomings. First and foremost is the fact that the results themselves to do not appear realistic. The convolutional autoencoders appear to smooth away upper and lower percentile temperatures and struggle to generalize. When tasked with just super-resolution, they fail to capture temperature gradients with elevation. This is even more difficult to assess because the authors fail to provide sufficient baselines to benchmark the machine learning approach. Furthermore, the presentation needs significant improvement for clarity. For these reasons, I recommend Major Revisions prior to consideration for publication in the Journal of European Geosciences Union (EGU). Below, I give recommendations to address the manuscript's shortcomings

Contents

1 Major Issues

1.1 Missing context in manuscript setup

The authors do a nice job framing the issue (lack of high-res GCM projection data) and why it is a problem. But two substantial pieces of context are neglected.

First, they insufficiently acknowledge and discuss other ML approaches to down scaling, in particular those which rely on deep, generative models like GANs [5, 3]. The authors should include appropriate citations to this work (and other ML based approaches for weather data super-resolution), but more importantly, justify the choice of an Autoencoder over an alternative machine learning approach. Second, the authors should acknowledge the other work of Autoencoders with meteorological data. Again, more critically, they need to better justify why autoencoders/convolutional networks should be used in a generative sense rather than for representation learning, which is more traditional and has shown clearer benefits [1, 2, 4].

Finally, the authors make the claim in lines 79-80 that convolutional methods generalize well. This requires citation.

1.2 Clarity in use of data

I would suggest the authors re-order the sections to discuss "Datasets" first. Providing context as to the details of the training (CMIP5) and target (ERA5) data will help non-domain readers more easily follow the logic of the authors proposed methodology in the Experimental setup.

The "Dataset Selection" itself should be condensed. It currently contains too much extraneous information about subsets of models. Much of the first several paragraphs should be cut or moved to the appendix as it will distract readers from the key takeaways needed about training and test data.

One of those pieces of information, the hourly sampling resolution is never defined. The authors state the time frame ERA5 data is available (hourly), and the data-range they look at (1975-2005), but they fail to clearly mention the temporal resolution they work at. Also, given the temporal auto-correlation of meteorological temperature data, the authors need to clarify the details of their training/validation split to ensure the autoencoders are trained objectively (the test split of taking the last year is clearly justified, however).

Additionally, the authors, with citations or other reasoning, should justify creating the upper and lower bounds of the min-max scaling (-50C to 50C) outside the min and max of their training data. This is going to force the data onto a narrower range than is necessary and I worry it limits the model performance.

1.3 Experimental Setup Issues

The diagrams of the experimental design (Figures 1 and 2) contradict each other. Figure 1 suggests that Autoencoder C takes in ERA 5 data (at low res.) but Figure 2 suggests it is used on CMIP5 data. Readers will find this confusing and it is important to clarify to interpret the results of the "RC" method. The nomenclature used will also cause confusion to readers. Having one approach called "AE" which is a very common abbreviation for Autoencoder is not ideal. Likewise, "RC" or Reconstruction is a bit inaccurate. The authors are not really reconstructing an image. They are generating a new one. This also gets to the issue that the authors are not using true Autoencoders, but rather Autoenocder-like models and that needs to be clarified in the manuscript. Additionally, the authors need to better justify architecture design choices including their decision to use MaxPooling over more common and modern MeanPooling or "striding" approaches.

1.4 Improper Baselines

It is difficult to fully assess the results given the lack of non-ML baselines.

Autoencoder C / RC approach is effectively just using non-linear dimensionality changes in the networks encoder and decoder for interpolation. The authors should include at least one conventional interpolation method for downscaling for comparison.

More importantly, the chosen comparisons (CMIP6, CMIP5, and CORDEX) all present issues. As noted by the authors, CMIP5 and CMIP6 are both GCMs with non-trivial biases. CORDEX does not encompass some critical regions including some of the most interesting topographical temperature gradients the authors are seeking to improve the representation of. What is missing is how this approach compares to a classical statistical down-scaling methodology. Solutions could include using the plate spline spatial interpolation from CCAFS-Climate or any other accepted method.

On the other hand, certain comparisons presently used seem to add little of value but serve to distract the reader from the results. In particular, it does not make sense to me to include CMIP6 data in this manuscript. It is not used to train the Autoencoders and the authors note in the "Dataset Analysis" section it has an exaggerated cold bias in this region so it is not useful as a measure of "ground truth". CMIP6 should be removed and replaced with results from classical interpolation and statistically downscaling.

1.5 Framing of Scientific Significance

I am concerned with the current framing of the results that this manuscript lacks the scientific significance in terms of novel results required of a submission to EGU. The results are mostly negative. The authors Autoencoders seem to under-perform at bias correction (where they under-predict high temperatures and over-predict low temperatures) and downscaling (where they do not properly capture geographic temperature gradients compared to a RCM). The paper would be better served as a narrative around the limits of convolutional neural networks. This would be useful to the community at a time when more and more teams are experimenting with neural network architectures and increasingly often convolutional neural networks.

But a second component of this requires the authors to think more about solutions. Particularly as it pertains to the Discussion/Conclusion. I do not believe extrapolating this approach, which has been demonstrated not to work, to other (less realistic) RCP scenarios would be productive. Another issue is if the authors are going to publish a section of the conclusion about the need for more/different methodologies ("Identifying methodologies to improve temporal resolution" and "Methodologies to increase precision over the ocean"), they need to specify some concrete suggestions or citations for specific ML approaches that could help correct the limitations of these shortcomings of the results. In "Addition of co-variants" authors need to specify what additional fields they think would be appropriate and why.

1.6 Presentation and Editing Issues

I am not convinced that as it stands the paper is up to EGU publication standards. In particular, the figures are lacking basic information including colorbars, proper captions, and visual editing. This needs to be improved before the paper can move forward. I will point out some specifics below

2 Technical Edits

(Line 49) [Missing citation to HighResMIP. Acronym not defined in advance.]

(Lines 84-86) [Clarify. What specifically about the geography of SEA will challenge convolutional methods?]

(Figures) [The first figure introduced in the paper is Figure 3 (line 100). The order of the figures should be changed so that they are referenced in order.]

(Line 145) [MSE not defined.]

(Figure 4) [Would recommend removing or putting in appendix.]

(Lines 187 and 228) [Year defined differently.]

(Figure 5) [Needs colorbar to be interpreted. Day should be specified in the caption.]

(Table 2 and elsewhere) [Should consider including the results for the interpolated (to CMIP5 resolution) ERA5 data.]

(Table 2 and elsewhere) [NIT. Use some form of brackets (e.g. [] or ()) for units rather than a "/" which looks like you are non-dimensionalizing the quantities.]

(Figure 6) [Colorbar needs label and units. Caption needs to specify the MSE between what and what. Consider log scale to better show differences over ocean and coastal plains for a more informative perspective.]

(Line 305) [Sentence missing a period.]

(Table 3 and Figure 8) [Move to the Appendix.]

(Figure S9) [Remove X and Y tick marks when you remove the labels. Reformat the subtitles.]

(Figure S10) [Remove Y ticks on RCP 4.5 panel.]

(Figures) [Give figures with multiple panels identifying labels (a, b, c, ect...).]

(Line 328) [Space after CMIP6.]

(Figure 13) [Colorbars needed.]

(Line 441) [Consider thickness of ocean "skin" as variable.]

References

[1] L. Denby. Discovering the importance of mesoscale cloud organization through unsupervised classification. *Geophysical Research Letters*, 47, 01 2020.

[2] T. Kurihana, E. Moyer, R. Willett, D. Gilton, and I. Foster. Data-driven cloud clustering via a rotationally invariant autoencoder. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–25, 2022.

[3] J. Liu, Y. Sun, K. Ren, Y. Zhao, K. Deng, and L. Wang. A spatial downscaling approach for windsat satellite sea surface wind based on generative adversarial networks and dual learning scheme. *Remote Sensing*, 14(3), 2022.

[4] H. Mangipudi, G. Mooers, M. Pritchard, T. Beucler, and S. Mandt. Analyzing high-resolution clouds and convection using multi-channel vaes, 2021.

[5] K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815, 2020.