Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-214-RC2, 2020

Geoscientific
Model Development
Discussions

Interactive comment on "Deep-learning based climate downscaling using the super-resolution method: a case study over the western US" *by* Xingying Huang

Anonymous Referee #2

Received and published: 28 December 2020

Overall comments

The author proposes a deep learning (DL)-based approach for super resolution (SR) in this article for climate downscaling. While the topic is an important one and the paper is reasonably well-written, the analysis is incomplete and insufficient for evaluating the DL model performance satisfactorily. The emphasis on the novelty of the DL approach is disproportionate, especially considering that the model architecture and training procedure are fairly standard. Furthermore, several recent DL-based SR approaches in the climate science literature are missing, which suggests that the author is not up-to-speed on the advancement of this field in the last couple of years. This is reflected in

the lack of breadth and depth of their analyses. As such, this paper is not ready for publication in GMD and will require more in-depth analysis and more critical evaluation of the results. Specific comments and suggestions are below.

Abstract

L14: "the" DL-based SR method? There are many, which one is being referred to here. L16: Only spatial SR is being done should be mentioned here in the abstract and in the body of the manuscript L17: It does not appear that there is anything new in this framework in terms of the DL architecture. Could the authors clearly state what is new and why they consider it "new"? L23-25: Some details can be given on why the DL model performance is competitive L26: It would help to be more specific about what types of coarse data this model can be used for, presumably it will not work well for all types of coarse climate data L28: In order to be applicable broadly, the DL model needs to be tested for generalization. Evidence of such analyses could be presented briefly in the abstract.

Body of the manuscript

L56: Some key references, including more advanced DL-based SR frameworks for climate downscaling, are missing: 1. Groenke et al. (arXiv:2008.04679Âă[cs.CV]) ClimAlign: Unsupervised statistical downscaling of climate variables via normalizing flows 2. Bano-Medina et al, https://doi.org/10.5194/gmd-2019-278 3. https://link.springer.com/article/10.1007/s00704-020-03098-3 4. https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.6769 5. Stengel et al., https://www.pnas.org/content/117/29/16805.short 6. Sha et al., https://journals.ametsoc.org/view/journals/apme/59/12/jamc-d-20-0057.1.xml 7. Sha et al., https://journals.ametsoc.org/view/journals/apme/aop/JAMC-D-20-0058.1/JAMC-D-20-0058.1.xml 8. Chen et al., https://arxiv.org/abs/2012.09700 9. Liu et al., https://doi.org/10.1145/3394486.3403366

L62-64: This statement is not true; see the refs above

L73-75: These and other advanced techniques have been used in other studies not referenced in this article. The so-called “up-to-date” and “cutting-edge” DL schemes used in this article are actually quite commonplace. In fact, current DL-based SR strategies in computer vision are more advanced than proposed in this article. c.f. Z. Wang, J. Chen and S. C. H. Hoi, "Deep Learning for Image Super-resolution: A Survey," inÂăIEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2020.2982166.

Sub-sections 2.1 and 2.2 can be compressed into a few key points. These are fairly standard DL methods, which is now quite well-adopted in the climate science community, hence getting into great depth is not warranted for a research article. Citing the key papers from the DL community and describing the chosen DL architecture briefly is sufficient. There are no novelties in these sections.

Table 1 is a helpful summary of the datasets used.

L207-208: Regridding using bilinear interpolation introduces a host of other challenges, in contrast to those from sophisticated super-resolution techniques. Hence the issues introduced by regridding need to be discussed.

L216: The training, validation and testing loss curves need to be shown to convince the reader that the model has converged and overfitting is not an issue âĂŤ not just the training loss curve. Further, the X axis of the loss curve is typically the number of epochs, not the number of iterations.

Section 3.1: The goal of SR is to capture fine scale details at daily or sub-daily timescales, hence showing the average over 10 years is not useful. As such, Figure 2 and Figure S3 are weak comparisons and do not show the power or limitations of the proposed DL SR model.

Figure 3 (left) appears to show that a simple bias correction of the ERA-I would shift the temporal evolution curve to be closer to the PRISM data. While Figure 3 (right) shows

that the seasonal cycle is almost identical for all models. What is the added value of a complex DL SR model here? Once again, it appears that testing annual averages is not very useful for an SR method, which aims to provide high spatial and temporal resolution. More useful tests include examining the spatial and temporal spectra; spatial and temporal coherent structures; extreme events (tails of the probability distributions of the physical variables downscaled); scatter plots or 2D PDFs to show false positives and false negatives.

Figure 4 is helpful to see differences at the daily timescale but should also be plotted on a log scale to be able to see the differences in the tails (extremes). Further, in depth examination of the extremes is needed to highlight the differences and plausible reasons for the differences.

Similar comments from Figure 3 for Figure 5 and 6 and S4 and S5. 10-year averages tend to smooth out errors in the SR process. Once again, it appears that testing annual averages is not very useful for an SR method, which aims to provide high spatial and temporal resolution. More useful tests include examining the spatial and temporal spectra; spatial and temporal coherent structures; extreme events (tails of the probability distributions of the physical variables downscaled); scatter plots or 2D PDFs to show false positives and false negatives.

Conclusions

L350: the novelties of the DL model shown here do not warrant this statement. L357-358: it is unclear what intuition was used in choosing the L1 vs L2 losses. L360: this is too strong a statement and I don’t think the results prove that, certainly not at fine spatial and temporal scales. L363: The role of elevation as an additional feature could be highlighted by training with and without this field to show the differences. This comparison would help make the usefulness of elevation more apparent. L365: Similarly for precipitation, ablation studies with and without the additional physical variables is needed to show their benefit. L374: Generalization studies are needed to make such strong statements. As such, the results shown do not warrant this statement.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-214, 2020.