Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2020-420-RC2, 2021

Comment on gmd-2020-420

Anonymous Referee #2

Referee comment on "Convolutional conditional neural processes for local climate downscaling" by Anna Vaughan et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-420-RC2, 2021

This paper presents a novel method for statistical downscaling leveraging the recent literature on neural processes, or more specifically, convolutional conditional neural processes (ConvCNPs). The authors present comprehensive experimental results for downscaling temperature and precipitation based on the well-established VALUE downscaling intercomparison framework.

Neural processes were undoubtedly a significant step forward in the field of probabilistic deep learning, and their value for the task of statistical downscaling is quite clear. Thus, the authors' work absolutely represents an important contribution to the literature and provides ample motivation for the continued investigation of ConvCNPs for downscaling, as well as related tasks in the future. The paper is also generally well written and the results are clearly presented.

However, despite its strengths, I do not think that this work is ready for publication as submitted. The literature review is insufficient and omits very relevant recent works in the field, e.g [2,4,7,8,9]. As a result, the authors significantly overstate their contributions on several occasions, in particular with regards to multi-site downscaling and generalization. Even worse, they misrepresent the current state of research on applications of deep learning in downscaling as somehow pessimistic, drawing a questionable contrast between the reported success of their method and the results of prior work. In fact, numerous authors have had success in applying deep learning to statistical downscaling tasks in recent years (see other references); thus the statement "...previous work [suggests] that little benefit is derived from applying neural network models to downscaling…" is, at best, misleading in the broader context of the literature.

Similarly, the repeated claim that "existing downscaling methods are unable to handle unseen locations" is not strictly true. It is typically possible, though not always effective, to train a downscaling model on one location and test it on another, provided that similar low resolution predictors are available. As noted in [1], convolutional neural networks are already well equipped for this when trained with data from multiple sites since they tend to learn more spatially invariant features. Transfer learning is also possible with other deep learning methods, as very recently demonstrated by [10]. ConvCNP may still generally be better, but the experiments presented by the authors do not show this. They only show ConvCNP is superior in comparison to a constructed Gaussian process interpolation baseline.

This brings me to the last major issue. The experimental analysis, while otherwise fairly rigorous and well designed, has one significant weakness: it does not, as far as I can tell, compare the authors' proposed ConvCNP method to any other deep learning methods recently proposed in the literature, including those that they cite such as the CNN architecture proposed by Baño-Medina et al. While this is not necessarily an absolute requirement for this paper to be a valuable contribution, it is at odds with how the authors seem to want to place their work in the context of current research. The authors suggest in sections 1, 2, and 6 that their approach is superior to existing deep learning methods. While this indeed may be true, it is not supported by their current results. I suggest that the authors should either add one or more existing deep-downscaling methods to their experiments, or re-frame their work and reduce the scope of their claims.

In summary, while the authors' work is undoubtedly a valuable contribution, the current presentation and framing within the context of the literature has significant problems. There is also a lack of much needed detail in the description of the ConvCNP model. The paper would benefit from a description which highlights step-by-step the similarities and differences of their architecture with the original ConvCNP architecture described by Gordon et al, as well as with other similar architectures like the ones described by Baño-Medina et al.

I will summarize my remaining technical comments by section. I look forward to seeing the authors' comments and revisions.

Section 1

- "Similar to what was mentioned previously, the statement "there has been debate as to whether deep learning methods provide improvement over traditional statistical techniques such as multiple linear regression" is perhaps overly pessimistic. There has been plenty of work now that clearly establishes the value of deep learning in statistical downscaling. Citing just two papers which happen to show somewhat mixed results in some cases comes off as a bit disingenuous to anyone who is familiar with the literature.

- "The second limitation common to existing downscaling models is that predictions can only be made at sites for which training data are available." Again, this is just simply not true. Nothing stops you from applying a model trained on location to another location where low-resolution predictors are available. How well it generalizes depends on the type of model used (e.g. a dense neural network will probably overfit) and geographic similarity. Where input or output grid interpolation is necessary, it is true that this can introduce additional error, but it is still certainly possible to generate "off grid" predictions.

- It would be helpful to provide precise definitions for "on-the-grid" and "off-the-grid" in the context of downscaling.

Section 2

- "ψ_MLP is a multi-layer perceptron, φ_c is a function parameterised as a neural network and CNN is a convolutional neural network." This sentence is a bit confusing. MLPs and CNNs are both types of neural networks. So you should be similarly more specific with what kind of neural network parameterizes φ_c.

- The definition of φ_c given in step 2 does not appear to be a neural network but rather a standard squared exponential kernel. The h term is produced by the CNN, but this was discussed separately. You should clarify if and where an additional neural network is used here (and what the optimized parameters are).

- The EQ summations are missing upper bounds. There must be some finite limit to the values of m,n computed here. Probably it is bounded by the size of the input (reanalysis) grid, but this should be stated explicitly as it has very significant implications on runtime complexity.

- It's not immediately clear why the EQ kernel is justified in this context. Is geographic proximity really that reliable a marker of similarity? Very distant regions can be similar and neighboring regions can be very different. Perhaps such distinctions are learned implicitly by the encoder. But regardless, a brief discussion of this question would be informative.

- "parameters at each target location θ" ☐ "parameters, θ, at each target location"

- "Formally, PP downscaling is an instance of a supervised learning problem to learn the function f in equation 1. Approaches to learning such a function have traditionally been split into two categories: …" The following statements make it sound like PP downscaling methods are always either neural networks or Bayesian, which is not true and probably not what you meant. Perhaps clarify whether or not you are talking about deep learning methods specifically (though, this would also not be true, as Bayesian DL methods have also been proposed [9]).

- In section 2.1.1, it's not explicit how the temporal dimension is handled. Is a separate distribution generated for each time step at each location? Or does each location get just one distribution from which samples are taken for every time step? Maybe specify a time index in the notation (at least initially).

- Maybe I missed it, but I don't think the ConvCNP and NP papers make any mention of using non-Guassian likelihoods like the Bernoulli-Gamma used here. It might be worth highlighting where this fits into the theoretical framework.

Sections 3 and 4

- Add units to the captions or y-axis of plots.

- MAE and bias tend to be tricky metrics for precipitation due to the heavy tails of the distribution and the prevalence of zero values on dry days. How is this handled here? Are they only calculated on days with precipitation?

- It would be nice to see the results for SDII and R01, space permitting.

- As mentioned by the other referee, KDE plots are inappropriate for the PIT in figure 10, as can be seen by non-flat appearance of the uniform distribution. Q-Q or P-P plots, or even just ROC curves, would be more illustrative. See [9] for examples of evaluating calibration on downscaling precipitation. I would advise against the idea of using a statistical test for uniformity on the grounds that statistical tests are generally somewhat uninformative and typically rely on arbitrary thresholds and questionable assumptions. A simple visualization is sufficient.

References

[1] Baño-Medina J, Gutiérrez JM. The importance of inductive bias in convolutional models for statistical downscaling. In Proceedings of the 9th International Workshop on Climate Informatics: CI 2019.

[2] Bürger G, Murdock TQ, Werner AT, Sobie SR, Cannon AJ. Downscaling extremes—An intercomparison of multiple statistical methods for present climate. Journal of Climate.

2012 Jun 15;25(12):4366-88.

[3] Cannon AJ. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. Computers & geosciences. 2011 Sep 1;37(9):1277-84.

[4] Groenke B, Madaus L, Monteleoni C. ClimAlign: Unsupervised statistical downscaling of climate variables via normalizing flows. In Proceedings of the 10th International Conference on Climate Informatics 2020 Sep 22 (pp. 60-66).

[5] Misra S, Sarkar S, Mitra P. Statistical downscaling of precipitation using long short-term memory recurrent neural networks. Theoretical and applied climatology. 2018 Nov;134(3):1179-96.

[6] Pan B, Hsu K, AghaKouchak A, Sorooshian S. Improving precipitation estimation using convolutional neural network. Water Resources Research. 2019 Mar; 55(3):2301-21.

[7] Singh A, White BL, Albert A. Downscaling numerical weather models with gans. In AGU Fall Meeting 2019 2019 Dec 12. AGU.

[8] Vandal T, Kodra E, Ganguly S, Michaelis A, Nemani R, Ganguly AR. DeepSD: Generating high resolution climate change projections through single image super-resolution. In Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining 2017 Aug 13 (pp. 1663-1672).

[9] Vandal T, Kodra E, Dy J, Ganguly S, Nemani R, Ganguly AR. Quantifying uncertainty in discrete-continuous and skewed data with Bayesian deep learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2018 Jul 19 (pp. 2377-2386).

[10] Wang F, Tian D, Lowe L, Kalin L, Lehrter J. Deep Learning for Daily Precipitation and Temperature Downscaling. Water Resources Research. 2021 Apr:e2020WR029308.