Using Multi-Head Attention Deep Neural Network for Bias Correction and Downscaling for Daily Rainfall Pattern of a Subtropical Island by Wang et al.

Main comments

In the study "Using Multi-Head Attention Deep Neural Network for Bias Correction and Downscaling for Daily Rainfall Pattern of a Subtropical Island" by Wang et al., a super resolution network is used to bias correct and downscale ERA5 reanalysis rainfall with 25 km horizontal spatial resolution to the observational TCCIP rainfall dataset with 5 km resolution over Taiwan. The performance of the method is extensively evaluated to a combination of quantile mapping and bilinear interpolation. Overall, the deep learning approach shows a promising performance in particular with respect to extreme event statistics.

The authors describe their method as being "tailored for future climate downscaling". In this context, it would be important to test whether the model can also preserve non-stationary trends that are expected in future warming scenarios? I would expect that such trends due to a warming of the atmosphere are already visible in the datasets used here.

Are there limitations with respect to extreme events not present in the training data, e.g. with return times much longer than the available reanalysis and observations?

The authors describe in section 2.3, that a main advantage of the deep learning approach is that the entire spatial domain can be processed at once instead of using grid cell-wise univariate quantile mapping. This should also be an advantage when correcting spatial patterns. Therefore, I believe it would be insightful to test correlations or power spectral densities computed over single daily fields.

Since a main focus of the study here is on the use and benefits of attention layers in the architecture, the network should be compared against a baseline architecture that is fully convolutional without attention.

Minor comments

L57: ".. the upscaling aspect.." is upsampling meant here? Upscaling would mean a coarsening of the resolution.

L134: How large are the 25km and 5km resolution fields in terms of pixel-dimensions?

L134: "Zeroes in" might not be the best wording here.

L166: ".. and encoder and an encoder“, the second one should be a decoder?

L176: Figure 2 instead of 3?

L207: Is the decoder or encoder trained here?

L235: Is empirical or parametric QM used?

L236: How many bins are used for QM? 15% bin width seems rather larger to me.

L461-463: As noted by the authors, stochasticity is crucial for modelling rainfall and uncertainty quantification with machine learning approaches. Since the network in this study includes dropout layers that can be used during inference for exactly this purpose, have you tried using them? It would be interesting to see the benefits.

L464: GANs are unsupervised and generally not part of reinforcement learning.

L469: To make this point, one should compare the network with attention used here against a fully convolutional network without attention layers.

Fig.1: Adding Fig.1c to Figure 3 would make the comparison much easier.

It would also greatly help the interpretation if all the result figures would include the model/dataset name as a subtitle.

Fig.3 Maybe add the color bar title as well?

Fig.7 Including a version of this figure without the log scale on the x-axis would improve the clarity of how the right tails are improved.

Fig.7 and 8. The color coding should be made consistent here to avoid confusion.

Tab.3. It would be important to see how correlation and RMSE are improved when computed between single days instead of mean precipitation only.