Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., author comment AC2
https://doi.org/10.5194/gmd-2022-163-AC2, 2023

Reply on RC2

Leroy J. Bird et al.

Author comment on "Deep learning for stochastic precipitation generation – deep SPG v1.0" by Leroy J. Bird et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-163-AC2, 2023

Reviewer comments are shown in bold.

The manuscript by Bird et al. presents a deep learning based stochastically generation of daily and hourly precipitation time series in New Zealand. The Authors used a neural network framework and probabilistic mixture distributions (i.e., gamma and GPD) to simulate the precipitation intensity at daily and hourly timescales. A list of statistics is shown to examine the synthetic time series and model performance against the observations and future climate projection. The Authors also present a non-stationary version of their deep learning model that incorporates contemporary/future changes in the precipitation traces through a temperature covariate. Overall, this manuscript contributes by adapting a neural network approach to stochastically develop multiple ensembles of precipitation at the regional scale. My comments/questions are listed below that are mostly related to the model structure, performance assessment, and precipitation temperature sensitivity parts:

We thank the review for taking the time to review the manuscript.

1) What was the reason for making the location parameter of GPD zero (L200)?

A GPD is typically used for modelling the tails of a distribution, however each component of the mixture model needs to be able to support the full range of potential precipitation depths when calculating the log-likelihood. For this reason, we set the location parameter to zero, despite this somewhat unusual use of a GPD distribution it did improve the representation of the extreme compared to only including Gamma distributions.

Also, were the parameters estimated at a seasonal scale

The model does have a seasonal signal, one of the input features to the NN is a sine and cosine term representing the seasonal signal. The NN learns how the distribution should change with respect to these terms.

or any other specific considerations applied to not mix up different rain-generating mechanisms into the same distribution?

There are no mechanism specific consideration, however the neural network is free to learn different mechanisms if they have an impact on the distribution of precipitation in the observation time series.

I presume a threshold or set of thresholds should also be considered to distinguish between the

mixture distributions (lower quantiles vs. heavy tail-like behavior); how did you set the

model to learn about these limits?

The distribution that is learned is inherently heavy-tailed so no specific distinction needs to be made. We uniformly sample from this distribution.

2) I was expecting that the developed stochastic precipitation generator model could also produce traces of precipitation reasonably higher or lower the observed time series across different ensemble members.

The SPG can produce values higher than observed in the training data set but not lower since it cannot produce precipitation values lower than 1mm / day for the daily SPG and 0.1 mm/hour for the hourly SPG.

However, it seems any values greater than the observed maximums (L249-252; Table 5) are replaced with equal or smaller magnitudes than the recorded maxima at daily and hourly timescales.

No, these are not observed maximums. These values are considerably higher than the previous events. They are simply upper bounds, which stop the SPG from producing

totally unrealistic precipitation depeths, which can very occotionally occur when uniformly sampling from the mixture model.

3) I would strongly suggest providing a few quantitative measures for the model comparison (e.g., percent bias, index of agreement, and so on) in addition to showing the model performance q-q plots in the "Stationary quality assessment" section for different precipitation characteristics, including moments, spells, extreme values of simulated traces. It is unclear how the stochastic model (e.g., different ensemble members) performs against the observations by only looking at these q-q plots.

We have now included metrics for dry days, moments, and extreme values comparing daily and hourly precipitation between observations and SPG simulations (Tables 6-9). We did not include an index of agreement as it would not be appropriate as the SPG doesn't provide a time series that is aligned with the observations. Furthermore, we thought including the metrics for the observations and the SPG is more informative than the percent bias.

4) How does the stochastic model preserve the year-to-year variability of seasonal/annual

precipitation variation?

The SPG has no capacity to preserve year-to-year variability, since it is conditioned only on the previous 8-day precipitation record and the autoregressive nature of the model. As such, we have no expectation that the SPG will capture the level of inter-annual variability observed in the real world as it has no way of learning of the variability on those time scales, and is not intended to be used for that purpose. We would expect year-to-year variability to be less than observed, given that the model includes no input features such as El Niño that drive longer-term variability. We have added text to highlight this caveat.

Can the Authors compare the intra-annual variation between the simulated traces and observed time series?

We have added the standard deviation of annual mean precipitation, for both the daily SPG and for observations, as a diagnostic in Table 8 for each of the four sites. Interestingly, there is no significant difference in the standard deviation in the annual mean precipitation derived from the dialy SPG and from the daily observations. This suggests that forced (as opposed to random) inter-annual variability in precipitation is rather small across these four sites in New Zealand - although an in-depth investigation of this conclusion is beyond the scope of this paper. For the hourly SPG, on the other hand, the variability in annual mean precipitation is underestimated at all locations. We have added some discussion of the annual variability to Section 4.5.

5) precipitation-temperature sensitivity: When I look at the P-T sensitivity plots,

it seems

negative scaling rates are calculated in many stations.

Correct.

Note that the Clausius–Clapeyron equation demonstrates a 'positive' relation between an increase in temperature and rainfall when the atmosphere is (nearly) saturated.

Yes, we are aware of this ,i.e., the Clausius–Clapeyron relationship suggests a +7%/K increase in the water carrying capacity of the atmosphere. We note, however, that this is different to a +7% increase in precipitation in that it is not only the water carrying capacity of the atmosphere that determines precipitation, but also climate-induced changes in the dynamics that may affect **where** the rain falls - in reality it is not unusual in many places for climate warming to induce decreases in light precipitation and increases in heavy precipitation.

As an example, it seems there is a strong 'hook' structure in the P-T relation across your stochastic precipitation generator/observation/model precipitation time series, and it should be taken into account before calculating the rates.

We agree and the hook was taken into account when calculating the rates, i.e., the hook structure is imposed onto the SPG.

6) "5.3 Post hoc addition of non-stationarity": It is unclear whether the established P-T

relationship is valid given Figures 16-17 and Comment#5 above.

We believe the figures are valid, given that the hook structure observed in Figure 16 was imposed onto the SPG when it was used to generate Figure 17.

Additional comments:

L117-124: I doubt using a single weather@home grid cell is a robust approach here in

general; how about using at least the four nearest grid cells to the site/station? This way,

the failure rate of encountering a 'bad' cell will statistically decrease fourfold.

We have used weather@home for several studies and, indeed, often use the average of the four nearest grid cells to the location of the observations as representative of the precipitation. However, for this study we only use the weather@home data to infer the sensitivity of the precipitation to a climate covariate, in this case the Southern Hemisphere land temperature anomaly - a field expected to be spatially more homogenous than the precipitation field itself.

Furthermore, it wasn't quite clear to us what the reviewer meant by a 'bad' cell. In our analyses of the weather@home data, we do sometimes find a cell with an occasional unrealistic precipitation value. However, we did screen the weather@home data to avoid such outliers. Our derivation of the sensitivity of the precipitation to such 'bad' cells is therefore mitigated.

L183-186: What computational settings were used to execute this neural network and

generate the traces? Please add some details about the logistics/computing configuration requirements.

The model was trained on a 24-core threadripper CPU. The model only took a few minutes to train on this machine. It could also be trained on a GPU. If the reviewer is asking about the python environment used for training the SPG, we have provided the conda environment .yaml file in the repository pointed to in the manuscript.

L196-203: Were these 12 parameters estimated for each site without considering any

seasonal influence on precipitation distributions? e.g., cold vs. warm season.

Yes, one of the input features to the neural network is a sine and cosine term representing the seasonal signal. The neural network learns how these parameters should change with respect to these terms.

L221-222: It is unclear what "better numerical stability" means here.

We have changed the wording, optimization using the likelihood requires calculating a joint product which quickly leads to floating point underflow for small probabilities, instead we optimize using log-likelihood which instead requires a sum, avoiding floating point

underflow.

L234-235: Is it causing a problem if the first eight days or 144 hours happen to be all

Zeros?

No. We anyway often observe dry period spells longer than 8 days and the model still exits those dry spells, creating dry-spell lengths consistent with the distribution of dry spells seen in reality (see Figure 7).

L312: What is your hypothesis on the hourly SPG that underestimates the seasonality in

the proportion of dry hours for Auckland?

We have no hypothesis but acknowledge that this underestimation could reduce with retraining of the SPG.

L321: Is this the entire section of the "4.5. Discussion" following the "4. Stationary quality

assessment"? I feel this "4.5" section does not add any new information more than what is

presented in "4.1" to "4.4" sections.

This is a good point made by the reviewer. We have deleted this section and ensured that any material covered by it now appears elsewhere.

In general, it seems some figures can be merged, or removed, as they do not add any

new information.

We carefully assessed and considered each figure and couldn't identify any that are surplus to the requirements of the paper.